

# **Book of Abstracts**

**18<sup>th</sup> Applied Stochastic Models and Data Analysis  
International Conference with  
Demographics Workshop**

**ASMDA2019**

*Editor*

**Christos H Skiadas**



**11 - 14 June 2019**

**Florence, Italy**

**Imprint**

**Book of Abstracts of the 18<sup>th</sup> Applied Stochastic Models and Data  
Analysis International Conference with the Demographics 2019  
Workshop**

**Florence, Italy: 11-14 June, 2019**

Published by: ISAST: International Society for the Advancement of  
Science and Technology.

Editor: Christos H Skiadas

e-book (PDF)

ISBN: 978-618-5180-32-4

## Preface

It is our pleasure to welcome the guests, participants and contributors to the International Conference (ASMDA 2019) on Applied Stochastic Models and Data Analysis and (DEMOGRAPHICS2019) Demographic Analysis and Research Workshop.

The main goal of the conference is to promote new methods and techniques for analyzing data, in fields like stochastic modeling, optimization techniques, statistical methods and inference, data mining and knowledge systems, computing-aided decision supports, neural networks, chaotic data analysis, demography and life table data analysis.

ASMDA Conference and DEMOGRAPHICS Workshop aim at bringing together people from both stochastic, data analysis and demography and health areas. Special attention is given to applications or to new theoretical results having potential of solving real life problems.

ASMDA 2019 and DEMOGRAPHICS 2019 focus in expanding the development of the theories, the methods and the empirical data and computer techniques, and the best theoretical achievements of the Applied Stochastic Models and Data Analysis field, bringing together various working groups for exchanging views and reporting research findings.

We thank all the contributors to the success of these events and especially the authors of this Abstracts Book. Special thanks to the Plenary, Keynote and Invited Speakers, the Session Organisers, the Scientific Committee, the ISAST Committee, Yiannis Dimotikalis, Aristeidis Meletiou, the Conference Secretary Eleni Molfesi, and all the members of the Secretariat.

May 2019

*Christos H. Skiadas*  
Conference Chair



## **ASMDA Conferences and Organizers**

- 1<sup>st</sup> ASMDA 1981 Brussels, Belgium. Jacques Janssen
- 2<sup>nd</sup> ASMDA 1983 Brussels, Belgium. Jacques Janssen
- 3<sup>rd</sup> ASMDA 1985 Brussels, Belgium. Jacques Janssen
- 4<sup>th</sup> ASMDA 1988 Nancy, France. J. Janssen and Jean-Marie Proth
- 5<sup>th</sup> ASMDA 1991 Granada, Spain. Mariano J. Valderrama
- 6<sup>th</sup> ASMDA 1993 Chania, Crete, Greece. Christos H Skiadas
- 7<sup>th</sup> ASMDA 1995 Dublin, Ireland. Sally McClean
- 8<sup>th</sup> ASMDA 1997 Anacapry, Italy. Carlo Lauro
- 9<sup>th</sup> ASMDA 1999 Lisbon, Portugal. Helena Bacelar-Nicolau
- 10<sup>th</sup> ASMDA 2001 Compiègne, France. Nikolaos Limnios
- 11<sup>th</sup> ASMDA 2005 Brest, France. Philippe Lenca
- 12<sup>th</sup> ASMDA 2007 Chania, Crete, Greece. Christos H Skiadas
- 13<sup>th</sup> ASMDA 2009 Vilnius, Lithuania. Leonidas Sakalauskas
- 14<sup>th</sup> ASMDA 2011 Rome, Italy. Raimondo Manca
- 15<sup>th</sup> ASMDA 2013 Mataró (Barcelona), Spain. Vladimir Zaiats
- 16<sup>th</sup> ASMDA 2015 Piraeus, Greece. Sotiris Bersimis
- 17<sup>th</sup> ASMDA 2017 London, UK. Christos H Skiadas

## SCIENTIFIC COMMITTEE

Jacques Janssen, Honorary Professor of Université Libre de Bruxelles, Honorary Chair  
 Alejandro Aguirre, El Colegio de México, México  
 Alexander Andronov, Transport and Telecom. Institute, Riga, Latvia  
 Vladimir Anisimov, Center for Design & Analysis, Amgen Inc., London, UK  
 Dimitrios Antzoulakos, University of Piraeus, Greece  
 Søren Asmussen, University of Aarhus, Denmark  
 Robert G. Aykroyd, University of Leeds, UK  
 Narayanaswamy Balakrishnan, McMaster University, Canada  
 Helena Bacelar-Nicolau, University of Lisbon, Portugal  
 Paolo Baldi, University of Rome "Tor Vergata", Italy  
 Vlad Stefan Barbu, University of Rouen, France  
 S. Bersimis, University of Piraeus, Greece  
 Henry W. Block, Department of Statistics, University of Pittsburgh, USA  
 James R. Bozeman, American University of Malta, Malta  
 Mark Brown, Department of Statistics, Columbia University, New York, NY  
 Ekaterina Bulinskaya, Moscow State University, Russia  
 Jorge Caiado, Centre Appl. Math., Econ., Techn. Univ. of Lisbon, Portugal  
 Enrico Canuto, Dipart. di Automatica e Informatica, Politec. di Torino, Italy  
 Mark Anthony Caruana, University of Malta, Valletta, Malta  
 Erhan Çinlar, Princeton University, USA  
 Maria Mercè Claramunt, Barcelona University, Spain  
 Marco Dall'Aglia, LUISS Rome, Italy  
 Guglielmo D'Amico, University of Chieti and Pescara, Italy  
 Pierre Devolder, Université Catholique de Louvain, Belgium  
 Giuseppe Di Biase, University of Chieti and Pescara, Italy  
 Yiannis Dimotikalis, Technological Educational Institute of Crete, Greece  
 Dimitris Emiris, University of Piraeus, Greece  
 N. Farmakis, Aristotle University of Thessaloniki, Greece  
 Lidia Z. Filus, Dept. of Mathematics, Northeastern Illinois University, USA  
 Jerzy K. Filus, Dept. of Math. and Computer Science, Oakton Community College, USA  
 Leonid Gavrilov, Center on Aging, NORC at the University of Chicago, USA  
 Natalia Gavrilova, Center on Aging, NORC at the University of Chicago, USA  
 A. Giovanis, Technological Educational Institute of Athens, Greece  
 Valerie Girardin, Université de Caen Basse Normandie, France  
 Joseph Glaz, University of Connecticut, USA  
 Maria Ivette Gomes, Lisbon University and CEAUL, Lisboa, Portugal  
 Gerard Govaert, Université de Technologie de Compiègne, France  
 Alain Guenoche, University of Marseille, France  
 Y. Guermeur, LORIA-CNRS, France  
 Montserrat Guillen, University of Barcelona, Spain  
 Steven Haberman, Cass Business School, City University, London, UK  
 Diem Ho, IBM Company  
 Emilia Di Lorenzo, University of Naples, Italy  
 Aglaia Kalamatianou, Panteion Univ. of Political Sciences, Athens, Greece  
 Udo Kamps, Inst. für Stat. und Wirtschaftsmath., RWTH Aachen, Germany  
 Alex Karagrigoriou, Department of Mathematics, University of the Aegean, Greece  
 A. Katsirikou, University of Piraeus, Greece  
 Włodzimierz Kłonowski, Lab. Biosign. An. Fund., Polish Acad of Sci, Poland  
 A. Kohatsu-Higa, Osaka University, Osaka, Japan  
 Tõnu Kollo, Institute of Mathematical Statistics, Tartu, Estonia  
 Krzysztof Kołowrocki, Depart. of Math., Gdynia Maritime Univ., Poland  
 Dimitrios G. Konstantinides, Dept. Stat. & Act. Sci., Univ. Aegean, Greece  
 Volodymyr Koroliuk, University of Kiev, Ukraine  
 Markos Koutras, University of Piraeus, Greece  
 Raman Kumar Agrawalla, Tata Consultancy Services, India  
 Yury A. Kutoyants, Lab. de Statistique et Processus, du Maine University, Le Mans, France  
 Stéphane Lallich, University of Lyon, France  
 Ludovic Lebart, CNRS and Telecom France  
 Claude Lefevre, Université Libre de Bruxelles, Belgium

Mei-Ling Ting Lee, University of Maryland, USA  
 Philippe Lenca, Telecom Bretagne, France  
 Nikolaos Limnios, Université de Technologie de Compiègne, France  
 Bo H. Lindqvist, Norwegian Institute of Technology, Norway  
 Brunero Liseo, University of Rome, Italy  
 Fabio Maccheroni, Università Bocconi, Italy  
 Claudio Macci, University of Rome "Tor Vergata", Italy  
 P. Mahanti, Dept. of Comp. Sci. and Appl. Statistics, Univ. of New Brunswick, Canada  
 Raimondo Manca, University of Rome "La Sapienza", Italy  
 Domenico Marinucci, University of Rome "Tor Vergata", Italy  
 Laszlo Markus, Eötvös Loránd University – Budapest, Hungary  
 Sally McClean, University of Ulster  
 Gilbert MacKenzie, University of Limerick, Ireland  
 Terry Mills, Bendigo Health and La Trobe University, Australia  
 Leda Minkova, Dept. of Prob., Oper. Res. and Stat. Univ. of Sofia, Bulgaria  
 Ilya Molchanov, University of Berne, Switzerland  
 Karl Mosler, University of Koeln, Germany  
 Amílcar Oliveira, UAb-Open University in Lisbon, Dept. of Sciences and Technology and CEAUL-University of Lisbon, Portugal  
 Teresa A Oliveira, UAb-Open University in Lisbon, Dept. of Sciences and Technology and CEAUL-University of Lisbon, Portugal  
 Annamaria Olivieri, University of Parma, Italy  
 Enzo Orsingher, University of Rome "La Sapienza", Italy  
 T. Papaioannou, Universities of Pireaus and Ioannina, Greece  
 Valentin Patilea, ENSAI, France  
 Mauro Piccioni, University of Rome "La Sapienza", Italy  
 Ermanno Pitacco, University of Trieste, Italy  
 Flavio Pressacco, University of Udine, Italy  
 Pere Puig, Dept of Math., Group of Math. Stat., Universitat Autònoma de Barcelona, Spain  
 Yosi Rinott, The Hebrew University of Jerusalem, Israel  
 Jean-Marie Robine, Head of the res. team Biodemography of Longevity and Vitality, INSERM U710, Montpellier, France  
 Leonidas Sakalauskas, Inst. of Math. and Informatics, Vilnius, Lithuania  
 Werner Sandmann, Dept. of Math., Clausthal Univ. of Tech., Germany  
 Gilbert Saporta, Conservatoire National des Arts et Métiers, Paris, France  
 Lino Sant, University of Malta, Valletta, Malta  
 José M. Sarabia, Department of Economics, University of Cantabria, Spain  
 Sergio Scarlatti, University of Rome "Tor Vergata", Italy  
 Hanspeter Schmidli, University of Cologne, Germany  
 Dmitrii Silvestrov, University of Stockholm, Sweden  
 P. Sirirangsi, Chulalongkorn University, Thailand  
 Christos H. Skiadas, ManLab, Technical University of Crete, Greece (Co-Chair)  
 Charilaos Skiadas, Hanover College, Indiana, USA  
 Dimitrios Sotiropoulos, Techn. Univ. of Crete, Chania, Greece  
 Fabio Spizzichino, University of Rome "La Sapienza", Italy  
 Gabriele Stabile, University of Rome "La Sapienza", Italy  
 Valeri Stefanov, The University of Western Australia  
 Anatoly Swishchuk, University of Calgary, Canada  
 R. Szekli, University of Wrocław, Poland  
 T. Takine, Osaka University, Japan  
 Andrea Tancredi, University of Rome "La Sapienza", Italy  
 P. Taylor, University of Melbourne, Australia  
 Cleon Tsimbos, University of Piræus, Greece  
 Mariano Valderrama, University of Granada, Spain  
 Panos Vassiliou, Department of Statistical Sciences, University College London, UK  
 Larry Wasserman, Carnegie Mellon University, USA  
 Wolfgang Wefelmeyer, Math. Institute, University of Cologne, Germany  
 Shelly Zacks, Binghamton University, State University of New York, USA  
 Vladimir Zaiats, Universitat de Vic, Spain  
 K. Zografos, Department of Mathematics, University of Ioannina, Greece

## **Plenary / Keynote / Invited Talks**

### **N. Balakrishnan**

Department of Mathematics and Statistics  
McMaster University, Hamilton, Ontario, Canada

### **Mark Brown**

Columbia University, USA

### **Robert J. Elliott**

Haskayne School of Business, University of Calgary, Canada and  
Centre for Applied Financial Studies, University of South  
Australia, Adelaide, Australia

### **Valérie Girardin**

Laboratoire de Mathématiques Nicolas Oresme,  
Université de Caen Normandie, Caen, France

### **Michael Katehakis**

Rutgers University, USA

### **Claude Lefèvre**

Université Libre de Bruxelles, Belgium

### **Sally McClean**

School of Computing and Information Engineering  
Ulster University, Coleraine, Northern Ireland

### **Isaac Meilijson**

Tel-Aviv University, Israel

### **Gilbert Saporta**

Conservatoire National des Arts et Métiers (CNAM), Paris, France

### **Dmitrii Silvestrov**

Stockholm University, Sweden



**Christos H Skiadas**

ManLab, Technical University of Crete, Greece

**Fabio Spizzichino**

Università degli Studi di ROMA "La Sapienza", Italy

**P.-C.G. VASSILIOU**

Department of Statistical Sciences,  
University College London, UK

**Vassilly Voinov**

KIMEP University, Almaty, Kazakhstan

<b>Contents</b>	<b>Page</b>
Preface	iii
ASMDA Conferences and Organizers	v
Scientific Committee	vi
Plenary / Keynote / Invited Talks	viii
Abstracts	1
Title Index	190
Author Index	212

# **BOOK OF ABSTRACTS**

**Applied Stochastic Models and Data Analysis  
ASMDA 2019 & DEMOGRAPHICS 2019**

## **Plenary and Keynote Talks**

### **Reliability Modelling and Assessment of a Heterogeneously Repaired System with Partially Relevant Recurrence Data**

**Narayanaswamy Balakrishnan**

*Department of Mathematics and Statistics, McMaster University, Hamilton,  
Ontario, Canada*

In this talk, I will first consider a reliability data to provide a basic motivation for the reliability problem considered in this work. Next, I will explain the stochastic modelling of the reliability problem and then describe the assessment methods for reliability. I will then revisit the data and illustrate the model and the assessment methods developed here. Finally, I will conclude the talk with some brief remarks and further suggestions!

### **Approximations with Error Bounds in Applied Probability Models: Exponential and Geometric Approximations**

**Mark Brown**

*Department of Statistics Columbia University USA*

Frequently in probability work simple approximations are sought for mathematically intractable probability distributions. Limit theorems often supply the approximating distribution, but what is really needed are error bounds for fixed  $n$  or  $t$ . In this talk I'll discuss some of my work over the years in error bounds for exponential and geometric distribution approximations. Points of interest include:

- 1) The waiting time for patterns in multinomial trials.

- 2) The first passage time to a set,  $A$ , for time reversible Markov chains.
- 3) The approximate exponentiality of geometric convolutions, with various applications.
- 4) The reliability of repairable systems.
- 5) Hazard function based bounds and inequalities.

## **New Filters for the Calibration of Regime Switching BETA Dynamics**

**Robert J. Elliott<sup>1</sup>, Carlton Osakwe<sup>2</sup>**

*<sup>(1)</sup>University of South Australia and University of Calgary*

*<sup>(2)</sup>Mount Royal University, Calgary*

In this paper we consider the estimation problem for reduced-form models that link the real economy to financial markets. Estimation is based on extending the work of Elliott and Krishnamurthy (1997, 1999) who derived new recursive filters to estimate parameters of a linear Gaussian, Kalman, filter. Some of the results were applied in Elliott and Hyndman (2007) to investigate commodity prices. This paper provides further extensions and also an application to calibrating a model for the beta of an industry, that is the process describing the sensitivity of an industrial sector's returns to broad market movements. The processes are scalar and hopefully, the new filtering methods easier to follow. In fact, the dynamics for the beta process of an industry are considered where the mean reversion level can take one of three values depending on whether the economy is in a 'good', 'medium' or 'poor' state. We assume the state of the economy is estimated using the growth rate of real GDP, and filtered estimates for the corresponding mean reversion level are used in a discrete time version of the beta dynamics. The beta process is estimated using the corresponding returns process and a new recursive filter is developed to estimate the mean reversion levels of the beta process.

## Entropy Rates of Markov Chains

**Valerie Girardin**

*Laboratoire de Mathématiques Nicolas Oresme, UMR6139,  
Université de Caen Normandie, Campus II, BP5186, 14032 Caen, France*

The definition by Shannon (1948) of entropy as a measure of uncertainty of a random phenomenon gave birth to information theory. Since then, many different generalized entropy functionals -- such as Renyi, Tsallis, Taneja, etc. -- have been defined for a better fit to complex systems.

The classical tool for studying random sequences is the entropy rate -- time averaged limit of marginal entropy of the sequence. The Shannon entropy rate of a countable Markov chain is the sum of the entropy of the transition probabilities weighted by the probability of occurrence of each state according to the stationary distribution of the chain; see Cover and Thomas (1991). Shannon and Renyi entropy rates are also functions of the Perron-Frobenius eigenvalue of some perturbation of the transition matrix; see Rached (2001, 2004).

Time averaged rates for all generalized entropy functionals but Shannon and Renyi are shown in Ciuperca, Girardin and Lhote (IEEE TIT, 2011) to be either infinite or zero. Nevertheless, averaging by some pertinent sequence -- induced by the asymptotic behavior of marginal entropy -- leads to meaningful generalized entropy rates in Girardin and Lhote (IEEE TIT, 2015), as soon as the random sequence satisfies a smoothness property. This quasi-power property is fulfilled by ergodic countable Markov chains under easy to check conditions. Closed-form expressions are thus obtained for the rates of Markov chains in Girardin, Lhote and Regnault (MCAP 2018).

**Keywords:** Entropy, Markov chains.

## Reinforcement Learning: Connections between MDPS and MAB problems

**Michael N. Katehakis**

*Rutgers University, Piscataway, NJ, USA*

In this talk we consider the basic reinforcement learning model dealing with adaptively controlling an unknown Markov Decision Process (MDP), in order to maximize the long-term expected average value. We show how a factored representation of the MDP problem allows it to be decoupled into a set of individual multi armed bandit (MAB) problems on a state by state basis. Additionally: i) we provide the construction of a simple UCB-type MDP policy, dramatically simplifying an earlier proof of its optimality, and ii) we discuss extensions to other MAB policies e.g., Thompson

Sampling. Talk based on joint work with Wesley Cowan, and Daniel Pirutinsky.

## Schur-Constant and Related Dependency Models

**Claude Lefèvre**

*Université Libre de Bruxelles  
Département de Mathématique  
B-1050 Bruxelles*

We consider the Schur-constant vectors in their continuous (usual) version and the discrete (less standard) version. Existing closed links with copulas and other dependency models are discussed. This leads us to examine and generalize the key properties of these models. As an illustration, we describe some applications of the Schur-constant models to insurance risk management.

This is a joint work with M. M. Claramunt and S. Loisel.

## Applied Stochastic Models: Theory vs Applications?

**Christos H Skiadas**

*ManLab, Technical University of Crete, Chania, Crete, Greece*

Following my experience by editing and publishing numerous studies presented in the last 27 years of my participation in ASMDA, I address this talk in the occasion of my 75 years. Theoretical issues look to overcome the applied part of Stochastic Models. However, many theoretical advances came after specific needs emerged in the real life. A difficult task is to collect and store data in a way to support a related theory. Another point is to develop a flexible theory to cope with the provided data. To this end we present a methodology combining theory and practice in Demography and especially in introducing stochastic modeling in human mortality and estimating the healthy life years lost. The results, after many years of work, support the importance of interconnection of theory and applications. Stochastic modeling is a quite strong tool for modeling real life applications, and real life provides enough variety in order to develop flexible applied stochastic models. Another challenge is related with transforming complicated stochastic models as to adapt to the data provided. And of course the interconnections with other scientific developments is extremely important. This is the case of Big Data modeling and artificial intelligence now at the core of scientific developments.

Few of the important parts of our work: When we used the first exit time theory to demography data sets an Inverse Gaussian was tested. Then

the data directed us to an Advanced form of the Inverse Gaussian not included in the theoretical tools. In this case the application directed us to extend the existing theory. Another case appeared when applied a first exit time theoretical model based on a first order approximation. To improve application, we tried a second order theoretical model with poor results. The solution came by developing a fractional approach methodology for computing first and second order derivatives.

Finally, the combination of both theory and applications is of particular importance. No theory is ready to cover any application and the applications are needed to improve or develop theory; if good data exist.

## **Applied and conceptual meaning of multivariate failure rates and load-sharing models**

**Fabio L. Spizzichino<sup>1</sup>**

*<sup>1</sup>University "La Sapienza", Rome, Italy*

The probability distribution of an absolutely continuous vector of lifetimes can be described in terms of the set of multivariate failure rate (m.c.h.r.) functions. In terms of such functions one can construct dependence models which are completely natural in several applied contexts and which do not admit any easy characterization in terms of the joint density function. This circumstance is in particular met in the situations of load sharing for the lifetimes of units which work as components in a system. It is assumed that, at any given age, the conditional hazard of any single component is actually influenced by the past failures of other components but it does not depend on the instants at which the failures had happened. In this talk, we describe main types of load-sharing models and we review old results and related applications. Then we show new results which also point out, for non-negative variables, the very meaning of the condition of absolute continuity and related implications in the analysis of stochastic precedence.

**Keywords:** Lifetimes of components in a systems, Absolutely continuous multivariate distributions for lifetimes, paradoxical aspects of stochastic precedence

## On the Theory and Applications of Nonhomogeneous Markov Set Systems

**P.-C.G. Vassiliou**

*Department of Statistical Sciences, University College London*

A more realistic way to describe a model is the use of intervals which contain the required values of the parameters. In practice we estimate the parameters from a set of data and it is natural that they will be in confidence intervals. In the present we study Non-Homogeneous Markov Systems (NHMS) processes for which the required basic parameters are in intervals. We call such processes Non-Homogeneous Markov Set Systems (NHMSS). Firstly, we study the set of the relative expected population structure and we prove with the help of Minkowski sums that under certain conditions of convexity of the intervals of the parameters the set is compact and convex.

## A note on serious arguments in favor of equality $P=NP$

**Vassilly Voinov**

*KIMEP University, Almaty, Republic of Kazakhstan*

A history of derivation of a polynomial in time algorithm for enumerating all existing nonnegative integer solutions of linear Diophantine equations, systems of equations and inequalities will be reminded.

Applications of the algorithm for solving: integer linear programs; 01, bounded and unbounded knapsacks; bounded and unbounded subset sum problems, and a problem of additive partitioning of natural numbers will be illustrated by numerous numerical instances.

A special attention will be devoted to solving (as majority of researchers think) NP-hard bin-packing and cutting stock problems. By the date only heuristic approaches for solving these two important for business problems are known. An application of the polynomial in time algorithm for solving bin-packing and cutting stock problems will be considered.

Arguments in favor of equality  $P=NP$  will be discussed.

**Keywords:** Formal Power Series, Linear Diophantine equations, Combinatorial Optimization, Integer Linear Programs, Partitions, Knapsacks and Subset Sums, Bin-packing and Cutting Stock Problems,  $P=NP$



## Invited Talks

### Sparse Correspondence Analysis

Ruiping Liu<sup>1</sup>, Ndeye Niang<sup>2</sup>, Gilbert Saporta<sup>3</sup>, Huiwen Wang<sup>4</sup>

<sup>1,4</sup>Beihang University, Beijing, China

<sup>2,3</sup>CNAM, Paris, France

Since the introduction of the lasso in regression, various sparse methods have been developed in an unsupervised context like sparsePCA which is a combination of feature selection and dimension reduction. Their interest is to simplify the interpretation of the pseudo principal components since each is expressed as a linear combination of only a small number of variables. The disadvantages lie on the one hand in the difficulty of choosing the number of non-zero coefficients in the absence of a criterion and on the other hand in the loss of orthogonality properties for the components and/or the loadings.

In this paper we are interested in sparse variants of correspondence analysis (CA) for large contingency tables like documents-terms matrices. We use the fact that CA is both a PCA (or a weighted SVD) and a canonical analysis, in order to develop column sparse CA and rows and columns doubly sparse CA.

**Keywords:** sparse methods, correspondence analysis, canonical analysis

### From Process Modelling to Process Mining: using Big Data for Improvement

Sally McClean

*School of Computing, Ulster University*

A process a series of tasks or steps taken in order to achieve a particular end, where well known examples are found in health, Internet of Things, transportation, smart grid, business, multi-player games, and fault prediction. Models of processes using tools such as Markov chain or Petri nets typically use a mathematical or symbolic model to provide a simplified representation of a system. Simulation can then use the model to imitate important aspects of the behaviour of the system and allow experimentation without having to disturb the real-life set-up.

In general, process mining aims to discover, monitor, and improve processes. This may include discovering the tasks within the overall processes, predicting future process trajectories, or identifying anomalous tasks or task sequences. Such process mining activities may build on

standard approaches to data mining problems such as classification, clustering, regression, association rule learning, and sequence mining or more recent approaches for Big Data, such as deep learning. However, if, or when, the structure of the process is known, model-based approaches can also be useful for incorporating structural process knowledge into the analysis and simplifying the problem.

Process mining thus unifies and builds on process model-driven approaches and classical data mining, using event logs, or other supplementary Big data, typically streamed, heterogeneous and distributed. It can be considered as a bridge between data mining and process modelling, providing a framework for design, an underpinning for process improvement and a scientific basis for decision making. Correctness/conformance and performance are among the important issues in the development of complex processes and systems, where process models are often used to assess such issues. Correctness can describe qualitative aspects of a system, such as liveness, safety, boundedness and fairness while compliance determines whether the observed process complies with the theoretical one. Performance describes the quantitative, dynamic, and time-dependent behaviour of systems, such as response time, system uptime, throughput or quality of experience.

We will discuss these concepts and approaches using a number of projects involving use-cases from healthcare, industry, networks, cloud and sensor technologies, computer games and pervasive computing.

## **Asymptotic Algorithms of Phase Space Reduction and Ergodic Theorems for Perturbed Semi-Markov Processes**

**Dmitrii Silvestrov**

*Department of Mathematics, Stockholm University, Stockholm, Sweden*

New asymptotic recurrent algorithms of phase space reduction and their applications to ergodic theorems for perturbed Markov chains and semi-Markov processes are presented. The classification of short, long and super-long time ergodic theorems for regularly and singularly perturbed Markov chains and semi-Markov processes is given as well as new limit theorems for hitting times and related functionals.

**Keywords:** Semi-Markov process, Perturbation, Phase space reduction, Ergodic theorem

## Talks for Invited and Contributed Sessions

### Structural Equation Modeling: Infant Mortality Rate in Egypt Application

**Fatma Abdelkhalek<sup>1</sup>, Marianna Bolla<sup>2</sup>**

*<sup>1</sup>Institute of Mathematics, Budapest University of Technology and Economics, Budapest, Hungary <sup>2</sup>Institute of Mathematics, Budapest University of Technology and Economics, Budapest, Hungary*

In social sciences, Structural Equation Modeling (SEM) is an important statistical approach for examining the causal relationships between variables. Since 1995 Egypt's infant mortality rate (IMR) has declined from about 50 deaths to infants under 1 year of age per 1000 live births to about 20 deaths in 2015 (World Bank 2015). In this paper we illustrate how SEM can be used to examine the factors that affect the IMR over time. We use data for five indicators: gross domestic product (GDP) per capita, current health expenditure as a percentage of the GDP, out of pocket health expenditure as a percentage of current health expenditures, 'Hepatitis B' immunizations, and the maternal mortality ratio, all available from the World Bank website. SEM results show the direct, indirect and total effects of each indicator on the IMR. SEM provides important sequential causal relationships that can help policy makers set program priorities.

**Keywords:** Structural Equations Modelling, Path Analysis, Recursive Regressions, Infant Mortality Rate

### A Topological Multiple Correspondence Analysis

**Rafik Abdesselam**

*University of Lyon, Lumière Lyon 2, COACTIS-ISH Laboratory, Lyon, France*

In this paper, we propose a new topological approach of data analysis which compares and classifies proximity measures for binary data. Based on the concept of neighborhood graph, this approach consists to select the "best" measure to analyze, understand and visualize the association between several categorical variables, the know problem of Multiple Correspondence Analysis (MCA). Similarity measures play an important role in many domains of data analysis. The results of any investigation into whether association exists between variables or any operation of clustering or classification of objects are strongly dependent on the

proximity measure chosen. The user has to select one measure among many existing ones. Yet, according to the notion of topological equivalence chosen, some measures are more or less equivalent. The concept of topological equivalence uses the basic notion of local neighborhood. We define the topological equivalence between two proximity measures, in the context of association between several categorical variables, through the topological structure induced by each measure. We compare proximity measures and propose a topological criterion for choosing the "best" association measure, adapted to the data considered, among some of the most used proximity measures for categorical data. The principle of the proposed approach is illustrated using a real data set with conventional proximity measures of literature for binary variables.

**Keywords:** Burt matrix; proximity measure; topological structure; neighborhood graph; adjacency matrix; topological equivalence and independence.

#### References

1. Abdesselam, R.: Selection of proximity measures for a Topological Correspondence Analysis. In a Book Series SMTDA-2018, 5th Stochastic Modeling Techniques and Data Analysis, International Conference, 12-15 June 2018, Chania, Crete, Greece, C.H. Skiadas (Ed), ISAST, 11-24, 2018.
2. Abdesselam, R.: Proximity measures in topological structure for discrimination. In a Book Series SMTDA-2014, 3rd Stochastic Modeling Techniques and Data Analysis, International Conference, Lisbon, Portugal, C.H. Skiadas (Ed), ISAST, 599-606, 2014.
3. Batagelj, V., Bren, M.: Comparing resemblance measures. In Journal of classification, 12, 73-90, 1995.
4. Demsar, J.: Statistical comparisons of classifiers over multiple data sets. The journal of Machine Learning Research, Vol. 7, 1-30, 2006.
5. Jaromczyk, J.-W. and Toussaint, G.-T.: Relative neighborhood graphs and their relatives. Proceedings of IEEE, 80, 9, 1502-1517, 1992.
6. Warrens, M. J.: Bounds of resemblance measures for binary (presence/absence) variables. In Journal of Classification, Springer, 25, 2, 195-208, 2008.
7. Zighed, D., Abdesselam, R., and Hadgu, A.: Topological comparisons of proximity measures. In the 16th PAKDD 2012 Conference. In P.-N. Tan et al., Eds. Part I, LNAI 7301, Springer-Verlag Berlin Heidelberg, 379-391, 2012.

## On PageRank Update in Evolving Tree Graphs

**Benard Abola<sup>1</sup>, Pitos Seleka Biganda<sup>1,2</sup>, Christopher Engström<sup>1</sup>,  
John Mango Magero<sup>3</sup>, Godwin Kakuba<sup>3</sup>, Sergei Silvestrov<sup>1</sup>**

<sup>1</sup>*Division of Applied Mathematics, The School of Education, Culture and Communication (UKK), Mälardalen University, Västerås, Sweden*

<sup>2</sup>*Department of Mathematics, College of Natural and Applied Sciences, University of Dar es Salaam, Dar es Salaam, Tanzania*

<sup>3</sup>*Department of Mathematics, School of Physical Sciences, Makerere University, Kampala, Uganda*

PageRank update refers to the process of computing new PageRank values after changes like addition or removal of links or vertices occurred in real life networks [2]. The purpose of updating is to avoid re-calculating the values from scratch. It is well known that handling nodes importance is problematic, particularly when links and nodes change [1]. In this talk, we are concerned with the problem of updating PageRank in changing tree graph. We present a few numerical experiments on a proposed algorithm that maintain level structures and update PageRank of evolving graph. Further, we will describe how to handle PageRank's update when cyclic components are formed via Jacobi-Chebyshev acceleration method and compare with the classical ones such as Jacobi and Power methods.

**Keywords:** PageRank, random walk, graph, Chebyshev acceleration

### References

- [1] A. N. Langville and C. D. Meyer. Google's PageRank and beyond: The science of search engine rankings. Princeton University Press, 2011.
- [2] C. Engström and S. Silvestrov. Calculating PageRank in a changing network with added or removed edges. In AIP Conference Proceedings (Vol. 1798, No. 1, p. 020052). AIP Publishing, 2017.

## Comparison of stability conditions for queueing systems with simultaneous service

**Larisa Afanaseva<sup>1</sup>, Svetlana Grishunina<sup>2</sup>**

<sup>1,2</sup>*Department of Probability Theory, Lomonosov Moscow State University, GSP-1, 1 Leninskiye Gory, Main Building, Moscow, Russia*

<sup>2</sup>*Moscow Institute of Electronics and Mathematics, National Research University Higher School of Economics, Moscow, Russia*

In this paper we study the stability conditions of the systems with  $m$  identical servers in which customers arrive according to a regenerative input flow  $X(t)$ . An arrived customer requires service from  $i$  servers simultaneously with probability  $\alpha_i$ .

A customer who arrives when the queue is empty begins service immediately if the number of servers he requires is available. Otherwise,

when a customer becomes first in a queue service begins immediately when the number of servers he requires is available. If a customer arrives to the system when the queue is not empty he goes to the end of the queue. We consider two cases: systems with independent service where service times by different occupied servers of a given customer are independent and systems with concurrent service where service times of a given customer are identical at all occupied servers. We compare stability conditions for the considered queueing systems for different number of servers. We perform a numerical analysis of dependence of the stability conditions upon service discipline and distribution of service times in the considered queueing systems for different number of servers.

Work is partially supported by Russian Foundation for Basic Research grant 17-01-00468.

**Keywords:** queueing theory, regenerative input flow, service discipline, simultaneous service

## Commute times and the effective resistances of random trees

Fahimnah Alawadhi

*Kuwait University*

The random walk on a graph  $G = (V, E)$  is a Markov chain defined on its set of vertices  $V$  and from vertex  $x$  it moves to a neighbor  $y$  chosen with uniform probability. The access time (hitting time)  $H_{xy}$  of a vertex  $y$  starting from a vertex  $x$  is defined to be the mean number of time units required to reach  $y$  for the first time. The commute time  $\tau_{xy}$  between  $x$  and  $y$  is the mean time units to go from  $x$  to  $y$  and then back to  $x$ .

That is,  $\tau_{xy} = H_{xy} + H_{yx} = \tau_{yx}$ . We study the commute and hitting times of simple random walks on spherically symmetric random trees in which every vertex of level  $n$  has out degree 1 with probability  $1 - q_n$  and 2 with probability  $q_n$ . Our argument relies on the link between the commute times and the effective resistances of the associated electric networks when one-unit resistance is assigned to each edge of the tree.

**Keywords:** random walk, Markov chain, hitting time

## Technical Efficiency of Public Healthcare Systems in Uttar Pradesh and Maharashtra: A Data Envelopment Analysis

**Cheryl Anandas**

*Tata Institute of Social Sciences.*

*21/3, A Wing, Prathamesh Apt, M.L. Camp, Matunga, Mumbai, Maharashtra, India*

Health is both a social and economic necessity. A basic level of health care needs to be assured to every citizen of the country to ensure physical and mental well-being of all the people. Technical Efficiency is defined as the effectiveness with which a given set of inputs is used to produce an output. Technical Efficiency addresses the issue of using given resources to maximum advantage; the productive ability to choose different combinations of resources to achieve maximum health benefit for a given cost. The Study has attempted to measure the technical efficiencies of health care facilities in all the districts of Uttar Pradesh and Maharashtra using data envelopment analysis. Data Envelopment Analysis is used to measure the efficiency of the Public Health Facilities. DEA is used to measure the district-wise efficiency of the healthcare system. DEA, as an analysis tool, has flexibility in handling multiple inputs and outputs, which makes it suitable for measuring the efficiency of hospitals that use multiple inputs to produce multiple outputs. The findings suggest that only 13 percentage of the healthcare facilities were acting as fully efficient facilities in the districts of Uttar Pradesh and 28 percentage in Maharashtra are fully efficient. There are various districts which have an efficiency score of more than 1 indicating the fact that the facility must increase its output levels to become efficient. The findings suggest that there exist individual districts who have efficiency less than 1 in some of the Healthcare System. To become efficient, these districts should be able to reduce their inputs without having to reduce their outputs regardless of the price of inputs. An important case is the district of Ratnagiri in Maharashtra; it is fully efficient in two out of the three health facilities regarding ANC, IFA tablets, delivery conducted and post-natal checkup. However, the prevalence of provision of all above mentioned services are very low. The results obtained in the study quantifies the efficiency of various services provided at public health sector at district level. However, it is not of much participation in some districts of Uttar Pradesh and Maharashtra

**Keywords:** technical efficiency, data envelopment analysis, Ante Natal Care, Iron Folic Acid tablets, Post Natal Checkup

## **Classification Methods for Healthcare System Costs in EU**

**A. Anastasiou, P. Hatzopoulos, A. Karagrigoriou, G. Mavridoglou**  
*University of Aegean, Dept of Stat and Actuarial-Financial Mathematics,  
 Samos, Greece*

The purpose of this work is to present and discuss Time Series Clustering Techniques and explore how they can be applied for the classification of the cost of national health systems in EU countries. We will describe various methods for Clustering and see the effect of similarity measures. One of the contributions of this work is the proposal of two new distance measures, called Causality Within Variables (CAWV) and Causality Between Variables (CABV) both of which are based on the well-known Granger Causality.

**Keywords:** Classification, Granger Causality, Similarity Measures

## **On demographic approach of the BGGM distribution parameters**

**Panagiotis Andreopoulos<sup>1</sup>, Alexandra Tragaki<sup>2</sup>, Fragkiskos G. Bersimis<sup>3</sup>, Maria Moutti<sup>4</sup>**

<sup>1,2</sup>*Department of Geography, Harokopio University, Athens, Greece,*

<sup>3</sup>*Department of Informatics and Telematics, Harokopio University, Athens, Greece,* <sup>4</sup>*Data scientist, Greece*

Analysis of the dynamics of human mortality during life is of great importance. Demographic comparisons between populations are expected to reveal differences in the causes of mortality that may be related to endogenous and exogenous factors. Identifying the formulation of health strategies and policies can be used to prevent or delay the aging process, reduce premature mortality, improve quality of life, and extend life span. The purpose of the study is to apply a new model of mathematical mortality (**BGGM** distribution) on data from Italy, for 114 years (1900 – 2013), in order to evaluate the proper adaptation of its new distribution and its values to its spatial and temporal differentiation. The application of the proposed approach is illustrated with the use of period death rates for the Italian population provided by the Human Mortality Database.

**Keywords:** Gompertz and Makeham functions, BGGM distribution, mortality dynamics, projections



## Shortest parts in Markov-modulated networks

A.M. Andronov<sup>1,2</sup>, I.M. Dalinger<sup>1</sup>, I.V. Yackiv<sup>2</sup>

<sup>1</sup>*Saint-Petersburg State University of Civil Aviation, Saint-Petersburg, Russia,*

<sup>1,2</sup>*Transport and Telecommunication Institute, Riga, Latvia*

Finite network is considered. Each network's arc has a constant length. The network operates in a random environment. The last is described by the finite irreducible continuous time Markov chain. Transition's speed along arcs depends on the state of the random environment. The algorithm for searching the shortest paths in the network is presented.

**Keywords:** Network, shortest path, random environment

## A comparison of graph centrality measures based on random walks and their computation.

Collins Anguzu<sup>1</sup>, Christopher Engström<sup>2</sup>, Sergei Silvestrov<sup>2</sup>

<sup>1</sup> *Department of Mathematics, School of Physical Sciences, Makerere University, Kampala, Uganda*

<sup>2</sup> *Division of Applied Mathematics, The School of Education, Culture and Communication (UKK), Mälardalen University, Västerås, Sweden*

When working with a network it is often of interest to locate the "most important" nodes in the network. A common way to do this is using some graph centrality measures. Since what constitutes an important node is different between different networks or even applications on the same network there is a large amount of different centrality measures proposed in the literature. Due to the large amount of different centrality measures proposed in different fields, there is also a large amount very similar or equivalent centrality measures in the sense that they give the same ranks. In this paper we will focus on centrality measures based on powers of the adjacency matrix or similar matrices and those based on random walk in order to show how some of these are related and can be calculated efficiently using the same or slightly altered algorithms.

**Keywords:** Graph, Graph centrality, Random walk, Adjacency matrix

## **Data-driven predictive modelling of enrolment associated processes to optimize clinical trial's operations**

**Vladimir Anisimov**

*Data Science, Centre for Design & Analysis, Uxbridge, London, UB8 1DH,  
United Kingdom*

Stochasticity and complex hierarchic structure of operational processes in large clinical trials require developing predictive analytical techniques for efficient modelling and forecasting trial operation. Predicting patient enrolment is among the major bottlenecks as uncertainties in enrolment substantially affect trial milestones and operational costs. The baseline analytic methodology for modelling enrolment using a Poisson-gamma model is developed by Anisimov & Fedorov (2005–2009). In the talk, several new developments and practical implementations in some areas are discussed. Predictive analytic modelling of enrolment on different levels (country/region) is considered. As usually there are only a few sites in a country and normal approximation, used in previous publications, cannot be applied, a new approach using the approximation of the enrolment process by a Poisson-gamma process with aggregated parameters is developed. The next area is predicting enrolment under some restrictions (low/upper country enrolment bounds). The optimal decision-making rules for data-driven interim re-projection and adjustment of enrolment are considered. The techniques for predictive modelling of some operational characteristics associated with enrolment, e.g. a number of non-enrolling or high enrolling sites, and forecasting performance in future time intervals using data-driven Bayesian re-estimation are also discussed. A novel analytic methodology for modelling jointly the processes of patients arriving for screening and randomized to a clinical trial is developed. A new approach to forecasting an optimal enrolment stopping time accounting for predictive number of randomized patients out of patients that are still in screening at interim time in order to minimize an excess of a sample size is proposed. The calculations are based on newly developed exact and approximation techniques and analytic/computational algorithms. Thus, Monte Carlo simulation is not required. The results are illustrated on several real case studies.

**Keywords:** Clinical trial, Modeling enrolment, Poisson-gamma model, Optimal design, Predictive modelling, Forecasting

## Robustness to outlying variables in PCA

**Arlette Antoni, Thierry Dhorne**

*Université de Bretagne Sud, CNRS UMR 6285, Lab-STICC, IUT de Vannes,  
BP 561, F-56017, Vannes cedex, France*

According to the duality framework inherent to PCA, the analyses in the observation space and in the variables space are fully consistent. This consistency nevertheless degrades when some constraints are added in one of both spaces. It is the case when dealing with robustness against outliers in the observational space. Minimization of a L1-norm criterion on the observations totally distorts the analysis in the variables space. It is therefore interesting to analyze PCA from a variables space framework. Paradoxically, although an outlying observation tends to influence the determination of factors and then to capture analysis, an outlying variable (here we mean a single variable uncorrelated with the others) tends to be pushed to the last factors with the risk of being excluded by the dimension reduction. While it is natural to reduce the influence of a dubious observation, it would be inadequate notably for end users of dimension reduction methods to let some variables totally disappear from the analysis. Surprisingly, this aspect is managed in PCA by providing a posteriori expertise. The aim of this work is to take into account robustness versus downplaying the interest of some variables by slightly modifying the criterion usually optimized in the variables space. In the first part of the presentation the problem of outlying variables is presented from a theoretical and practical point of view. Then a modification of the inertia criterion used in PCA is proposed to improve the influence of any variable. In a second part, it is shown that classical optimizers can give efficient solution to the maximization problem and can lead to interesting analysis. Finally some algorithmical modifications are suggested both to improve precision and to mix the method under discussion with classical PCA. Throughout the presentation a classical data set used in dimension reduction is worked out. Functions are made available in R, Python and/or Julia.

**Keywords:** PCA, Outlying variables

## Floods: Statistics, Analysis, Regulation

**Valery Antonov, Roman Davydov**

*Department of Mathematics, Peter the Great St. Petersburg Polytechnic University, Saint-Petersburg, Russia*

Floods constitute a third of natural disasters and economic losses, and often lead to human victims. It is predicted that these phenomena will become more frequent in the future due to population growth, urbanization, land depletion and climate change. In this regard, the increasing importance is attached to the creation of adequate models for the study of these complex processes. Among them, a special place is occupied by mathematical models.

Mathematical models are used to solve the following problems:

- modeling of floods on rivers and their tributaries;
- development of effective measures for flood protection;
- assessment of possible damage;
- analysis of water consumption in various weather conditions.

Mathematical models can be classified as stochastic, deterministic and chaotic. This paper discusses the creation and application of these mathematical models. The focus is on models based on stochastic analysis. The basis of stochastic models is frequency analysis, which makes it possible to determine the likelihood that flooding will happen or not happen in the near term.

**Keywords:** Floods, Mathematical Models, Frequency Analysis

## Fractal analysis of nanostructured material objects

**Valery Antonov<sup>1</sup>, Anatoly Kovalenko<sup>2,1</sup>**

*<sup>1</sup>Mathematics Department, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

*<sup>2</sup>Physics Department, Joffe Physical-Technical Institution Russian Academy of Science, St. Petersburg, Russia*

The method of the fractal analysis is based on the general idea of Mandelbrot of the difference between the topological Euclidean dimension of such non-smooth objects and the geometric dimension with the invariant metric-statistical self-similarity of their different-scale structures. Mathematically, this manifests itself in the form of the dependence  $M \sim \varepsilon^D$  between the rate of increase in the number of structural elements  $M$  under

consideration and the increase in the spatial scale interval  $\varepsilon$  of their consideration with a fractional index of fractal dimension  $D$ . A similar pattern takes place when considering the occurrence of fractal processes of deterministic chaos in time. To characterize the ordered models of regular mathematical fractals, a single  $D$  value is sufficient. However, an adequate description of real disordered (nonuniform) natural fractals and many irregular model structures along with metric characteristics require determination of their statistical properties reflected by the full spectrum of fractal dimensions using a multifractal formalism. Establishing the characteristics of these dependencies allows us linking the indicators of structural and phase nonuniformity in the development of new materials with changes in their physicochemical properties. The paper presents the developed gradient-pixel method of fractal analysis and results of multifractal characterization of nanostructured materials with a high proportion of non-autonomous phases obtained from micrographs of their surface chips with high-resolution scanning microscopes. In comparison with the fractal dimension of the Sierpinski carpet as a classic regular monofractal computed on the outlined basis, quite accurately coinciding with the known analytical value, the resulting spectrum of fractal dimensions of the synthesized chemical-catalytic and thermoelectric nanomaterials indicates the multifractal nature of their structural and phase nonuniformity according to the Rényi generalized equation.

**Keywords:** Fractal analysis, Gradient-pixel method, Spectrum of multifractal dimensions, Nanostructured materials, SEM micrographs of chips

## Kernel SVM Distance Based Control Chart for Statistical Process Monitoring

**Anastasios Apsemidis, Stelios Psarakis**

*Department of Statistics, Athens University of Economics and Business,  
Athens, Greece*

Traditional statistical process control methods are based on a parametric model, which is used to differentiate in and out of control data and attain process stability. In the present article, we propose the use of a non-parametric method that transforms the classical process control problem into a classification problem utilizing the Support Vector Machines theory. The chart monitors the probability of the process being out of control eliciting information from a moving window and the position of new

observations in relation to a decision boundary. The proposed method is tested via simulations under different scenarios and a real data example is also included to verify the effectiveness of the new chart.

**Keywords:** Kernel methods, support vector machines, statistical process monitoring, multivariate control chart, machine learning

## Modeling private preparedness behavior against flood hazards

**Pedro Araujo<sup>1</sup>, Gilvan Guedes<sup>2</sup>, Rosangela Loschi<sup>3</sup>**

*<sup>1,2</sup>Department of Statistics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, <sup>2</sup>Demography Department, Av., UFMG, Belo Horizonte, Brazil*

Floods are one of the major causes of health and economic losses among natural disasters worldwide. Preparedness behavior against flood hazards however is low in many parts of the world, including regions with high levels of environmental awareness and economic development. The Protective Action Decision Model (PADM) is a well-established conceptual framework proposed in social psychology, predicting that the intention to adopt measures against floods is positively associated with the perceived effectiveness of these measures and negatively related to their direct and indirect (opportunity) costs. Despite initiatives to measure effectiveness and opportunity costs of private measures against hazards exist, results are mixed. We argue that the mixed findings are the result of two processes: 1) risk aversion is key in any private insurance model, but the studies using PADM do not include this variable (omission bias), and 2) the effects of effectiveness and costs are latent variables that indirectly measure personal traits. This study investigates if the hypotheses supporting PADM are satisfied in a study involving individuals under risk of river floods in Brazil. Our model improves previous efforts in many ways: 1) it is based on a probabilistic sample, with 1164 individuals interviewed in a city with a large share of the population under risk of river floods; 2) it introduces a hierarchical Bayesian logistic model relating the probability of adopting protective measures against floods to covariates directly measured from individuals, as well as to the latent covariates representing risk-aversion and perceptions about the effectiveness (PE) and the opportunity cost (PCO) of those measures; 3) it measures PE and PCO through item response theory (IRT) models, allowing us to appropriately quantify the uncertainty inherent to such quantities; 4) it includes a random

effect reflecting unmeasured individuals features to correlate the individual responses to the different protective measures considered in our study. We found that the effect of PCO is small and negative for protective measures in contrast to the high and positive effect of PE, net of risk aversion and income as predicted by PADM.

**Keywords:** Protection Action Decision Model, Item Response Theory, Random Effect, Risk Aversion, Brazil, flood hazards

## Inflation Rates Indicators and their Properties

**Josef Arlt, Markéta Arltová**

*Department of Statistics and Probability, University of Economics Prague*

Inflation as an important feature of certain economic area is in the relationship with the rate of economic growth, it affects the value of money and reflects the rate of economic stability. It plays irreplaceable role for example in the monetary policy; in the valorization of wages, pensions and social benefits, and so on. Therefore, the measures of inflation are extremely important. The source of information on price levels is the consumer price index, which measures the price development of the basket of goods and services consumed during the base period. The consumer price indexes of different countries are typified by the increasing trends and cyclical and seasonal movements, which are sometimes not clearly seen. The inflation rates are calculated as a growth rates of the consumer price index. They give the information on its dynamics and it can be calculated in several ways. Each inflation rate has particular statistical properties. The monthly inflation rate is defined as the month-over-month growth rate of the consumer price index. The monthly inflation rates of different countries are close to the stationary time series. They contain the cyclical components which may not be visible because they are usually covered by the pronounced seasonal and irregular components. The annualized inflation rate is a hypothetical measure which informs on the potential inflation rate per year under the conditions of the given monthly inflation rate. This time series thus copies the development of the monthly inflation rate, but at a different level. The annual inflation rate is defined as the year-over-year growth rate of the consumer price index, which is, in fact, the seasonal difference of the logarithm of the consumer price index. It can be computed as a one-sided

simple moving average of twelve annualized inflation rates, which is the smoothing filter with special properties: (A) There is a lag time of 5.5 months between a turning points of the original time series and the filtered time series. (B) The moving average removes some part of the seasonality and noise from the original time series and leaves the cyclic components. This transformation leads to the higher persistence in the comparison with the monthly and annualized inflation rates. The persistence is related to the observed cycles in the annual inflation rate. It is shown that the cycles are spurious, therefore the use of the annual inflation rate in practice is problematic.

**Keywords:** Inflation, time series, cycles

## A new simplex distribution allowing for positive Covariances

**Roberto Ascari<sup>1</sup>, Sonia Migliorati<sup>2</sup>, Andrea Ongaro<sup>2</sup>**

<sup>1</sup> *Department of Statistics and Quantitative Methods (DISMEQ) Via Bicocca degli Arcimboldi, 8, Università di Milano-Bicocca, Milano, Italy* <sup>2</sup>*Department of Economics, Management and Statistics (DEMS) P.zza dell'Ateneo Nuovo, 1, Università di Milano-Bicocca, Milano, Italy*

Vectors of proportions arise in a great variety of fields: chemistry, economics, medicine, sociology and many others. Supposing that a whole can be split into  $D$  mutually exclusive and exhaustive categories, vectors describing the percentage of each category on the total are referred to as compositional data. The latter are subject to an unit-sum constraint and thus their domain is the  $D$ -part simplex. A very popular distribution defined on the simplex is the Dirichlet one. This distribution, despite its several mathematical properties, is poorly parametrized and, therefore, it cannot model many dependence patterns. Some authors proposed alternatives to the Dirichlet, looking for more flexible distributions which still retain some relevant properties for compositional data. Among these is the Flexible Dirichlet (FD), introduced by Ongaro and Migliorati [1], which generalizes the Dirichlet distribution, that is included as an inner point. Thanks to its mixture structure with  $D$  components, it exhibits a more suitable modelization of the covariance matrix. Despite its greater flexibility, the FD lacks in allowing for positive covariances, which are plausible in many applications. The aim of this contribution is to present a



further generalization of the Dirichlet, called Double Flexible Dirichlet (DFD), that takes advantage of a finite mixture structure similar to the one of the FD (depending on  $D(D + 1)/2$  mixture components) and enables positive covariances. Some theoretical results are shown and an estimation procedure based on the EM algorithm is proposed, including an ad hoc initialization strategy. A simulation study aimed at evaluating the performance of the EM algorithm under several parameter configurations is included.

**Keywords:** Compositional Data, Dirichlet Mixture, EM algorithm.

#### References

1. A. Ongaro and S. Migliorati. A generalization of the dirichlet distribution. *Journal of Multivariate Analysis*, 114(1):412-426, 2013

## Application of meta-heuristic optimization approaches in operational planning for clinical trials

Matthew Austin<sup>1</sup>, Stephen Gormley<sup>2</sup>, Vladimir Anisimov<sup>2</sup>

<sup>1</sup>Center for Design & Analysis, Amgen Inc, Thousand Oaks, CA, USA

<sup>2</sup>Center for Design & Analysis, Amgen Ltd, Uxbridge, London, UK

At the initial stage of the enrolment design in multi-country clinical trials, our goal is to find an optimal geographic distribution of clinical sites (allocation in countries) that meets several requirements: (a) site allocation can achieve target enrollment time goal at a given probability (e.g. 80%), (b) site allocation represents minimum trial cost, (c) number of sites in countries must be within pre-specified country-specific upper and lower bounds, and (d) number of patients recruited in a country is not more than a pre-determined country maximum number. Using the enrollment process models developed in Anisimov & Fedorov (2007, 2011), we can derive the analytic/computational expressions for the probability of achieving the enrollment target given the country specific constraints on the number of sites and the maximum number of patients. Using this approach allows to determine an optimal trial design with the minimal cost given country's and sites/patients restrictions. Criterion of optimality. Find an allocation of sites  $\{N_1, \dots, N_J\}$  in  $J$  countries that minimizes  $\text{TrialCost}(N_1 \dots N_J) = \sum_j C_j N_j + C_{\text{fixed}}$  given the probability to complete enrollment by the planned date meets a specified probability  $\Pr(\text{EnrollmentTime}(N_1, \dots, N_J, \text{Cap}_j) \leq T_{\text{plan}}) \geq P_{\text{plan}}$  where  $L_j \leq N_j \leq U_j, j=1, \dots, J$  and enrollment in country  $j$  is restricted by  $\text{Cap}_j$ .

Combinatoric issues. An exhaustive search of the optimal allocation would require  $\prod_j (U_j - L_j + 1)$  evaluations of the objective function. As the number of countries in the trial is normally  $\geq 15$ , then in a 15-country scenario where each country could contribute up to 5 sites, an exhaustive search would require 615 evaluations, which is not achievable in the real time. We have approached this problem using meta-heuristic global optimization techniques. We will share our findings and discuss the strengths and weaknesses of this approach.

**Keywords:** Optimal enrollment design, Poisson-gamma model, Combinatorial optimization, Genetic algorithm, Differential evolution

### Increasing efficiency in the EBT algorithm

**Tin Nwe Aye<sup>1,2</sup>, Linus Carlsson<sup>3</sup>**

<sup>1</sup>Department of Mathematics, University of Mandalay, Myanmar <sup>2</sup>Division of Applied Mathematics, Mälardalens University, Västerås, Sweden <sup>3</sup>Division of Applied Mathematics, Mälardalens University, Sweden

The Escalator Boxcar Train (EBT) is a commonly used method for solving physiologically structured population models. The main goal of this paper is to overcome computational disadvantageous of the EBT method. We prove convergence, for a general class of EBT models in which we modify the original EBT formulation, allowing merging of cohorts. We show that this modified EBT method induces a bounded number of cohorts, independent of the number of time steps. This in turn, improve the numerical algorithm from polynomial to linear time. An EBT simulation of the Daphnia model is used as an illustration of these findings.

**Keywords:** Escalator Boxcar Train, structured population models, merging.

### Clustering of multiple lifestyle risk factors and health-related quality of life in Korean population

**Younghwa Baek, Kyungsik Jung, Hoseok Kim**

*Korea institute of oriental medicine*

Background. Several studies have shown elevated risk of cardiovascular disease (CVD) associated with certain lifestyle habits. A combination of

two or more risk factors was closely associated with a higher increase risk of CVD than can be expected on the basis of the sum of the individual effects. The aim of study was to examine the clustering pattern of three lifestyle risk factors, focusing on sleep, physical activity, and eating habits, and to explore the relation with the health related of quality of life (HRQOL) among Korean population.

**Methods.** Data on lifestyle risk factors (poor sleep quality, low physical activity, poor eating habits), sociodemographic characteristics, and HRQOL using SF-12(version 2) were obtained from 5,221 men and women  $\geq 18$  years of age who participated in this study. We calculated the ratio of the observed to expected (O/E) prevalence for the 8 different combinations and the prevalence odds ratios (POR) of three lifestyle risk factors. Logistic regression analysis adjusted for sex and age was used to assess whether the clustering of multiple risk factors was independently associated with HRQOL.

**Results.** The three lifestyle risk factors tended to cluster in specific multiple combinations. Poor sleep quality and low physical activity was clustered (POR: 1.32 for men), Poor sleep quality and poor eating habits were clustered (POR: 1.46 for men, 1.34 for women), and low physical activity and poor eating habits were also clustered (POR: 1.48 for men). The increased lifestyle risk factors clustering was significantly decreased with physical and mental HRQOL (Odds ratio for low physical HRQOL = 1.3, 1.6, and 2.4; Odds ratio for low mental HRQOL= 1.6, 2.1, and 2.9, by having 1,2, and 3 lifestyle risk factors, respectively)

**Conclusions.** These findings suggest that common lifestyle risk factors cluster among adult subjects. We might get help through knowledge on clustering pattern of lifestyle risk factors for more effective intervention in public healthcare system.

**Keywords:** Lifestyle, Clustering, Health related quality of life medicine

## Multiple outliers identification in linear regression

Vilijandas Bagdonavičius<sup>1</sup>, Linas Petkevičius<sup>2</sup>

<sup>1,2</sup>Vilnius University, Vilnius, Lithuania

After strict definition of outliers, a new very competitive method for multiple outliers identification in linear regression models is proposed.

The hypothesis of the absence of outliers is rejected if the proposed test statistic takes large value. Asymptotic distribution of the test statistic is studied and approximations for the critical values of this statistic are given.

Classification rules of observations to outliers and non-outliers are strictly defined. The new method is compared with well-known outlier identification methods by simulation. The masking and swamping values and also other characteristics were computed for various sample sizes and various outlier generation schemes. Comparative analysis is also done using numerous real data examples.

**Keywords:** Outlier identification, Linear regression, Multiple outliers, Outlier region, Robust estimators

### Residual based goodness-of-fit tests for linear regression models

Vilijandas Bagdonavičius<sup>1</sup>, Rūta Levulienė<sup>2</sup>

<sup>1</sup>Vilnius university, Naugarduko 24, LT-03225 Vilnius, <sup>2</sup>Vilnius university, Naugarduko 24, LT-03225 Vilnius, Lithuania

We consider a general parametric linear regression model  $Y = \beta_0 + \beta_1 z_1 + \dots + \beta_m z_m + \sigma \varepsilon$ , where  $\beta_0, \dots, \beta_m$ ,  $\sigma$  are unknown parameters,  $z_1, \dots, z_m$  are observable covariates,  $\varepsilon$  is a non-observable zero mean (or zero median if the mean does not exist) absolutely continuous random variable with a cumulative distribution function  $F$ . Our purpose is to test the hypotheses  $H: F = F_0$  for specified functions  $F_0$ . As examples, goodness-of-fit for normal, Cauchy, extreme value, Weibull, continuous logistic and loglogistic regression models are considered. We propose a class of simple and powerful tests based on residuals. For finite samples, the critical values of the test statistics were found by simulation. The asymptotic law and approximations of the critical values of the test statistics were found, too. For various sample sizes and alternatives the power of the tests was investigated by simulation.

**Keywords:** Goodness-of-fit, linear models, Cauchy regression, normal regression

## Structure of the particle field for a branching random walk with a critical branching process at every point

**Daria Balashova<sup>1</sup>, Stanislav Molchanov<sup>2</sup>, Elena Yarovaya<sup>1</sup>**

<sup>1</sup>*Department of Probability Theory, Faculty of Mechanics and Mathematics,  
Lomonosov Moscow State University, Moscow, Russia*

<sup>2</sup>*Department of Mathematics and Statistics, University of North Carolina at  
Charlotte, NC, USA; National Research University, Higher School of Economics,  
Moscow, Russia*

We consider a continuous-time branching random walk (BRW) on the multidimensional lattice  $\mathbb{Z}^d$ ,  $d \geq 1$ . We assume that at the initial moment there is a particle at each point of the lattice. The spatio-temporal evolution of the random field of the particles includes the symmetric walk with a finite variance of jumps over  $\mathbb{Z}^d$  and birth-death processes at every lattice point. The intensities of birth and death of particles are assumed to be equal at each point of the lattice. There is no interaction between particles. All descendants of the particle located at the initial moment of the time at the point  $x$  will be called the subpopulation of this particle and denoted by  $n_x(t)$ . We obtain for such BRW the following situation: the majority of the subpopulations  $n_x(t)$  are vanishing on the large time interval, but the remaining part of subpopulations, which denoted by  $n_{x_i}$  will have the order  $O(t)$  at least at the level of the first moment. We prove that the vanishing of the majority of  $n_x(t)$  doesn't have an effect on the convergence to a steady state of a particle system for a transient random walk and leads to clusterization of particles for a recurrent random walk. Moreover, the process of clusterization with gaps between clusters of particles is simulated. The authors<sup>1</sup> were supported by the Russian Foundation for Basic Research, project No. 17-01-468. The author<sup>2</sup> was supported by the Russian Science Foundation, project No. 17-11-01098.

**Keywords:** Branching process, Branching random walks, Population dynamics, Large deviations

## Detecting long term and abrupt changes in hydrological processes

**Dominika Ballová, Adam Šeliga**

*Faculty of Civil Engineering, Slovak University of Technology, Bratislava*

Detection of changes plays an important role in various fields of study. In hydrological processes it is crucial to detect changes, since it can help

preventing or at least prepare for extreme events like floods and drouth. Usually we can observe two kind of changes; long term changes displayed in the data as a trend and sudden switch in parameters of the distribution represented by change-points. Various methods for observing changes has been developed. One can be interested in detecting the presence of one significant step change in the series. By increasing the number of observations of the series, multiple change-points detection can be a useful tool. Also determining long lasting increasing or decreasing trend in the series is a frequently studied issue. In this paper, changes in the development of hydrological time series of main Slovak rivers were detected. Since the data follow non normal distribution, we obtained our results by means of nonparametric methods. Significant trends in the series were detected by applying Mann-Kendall test, Spearman's Rho test and Cox-Stuart test. Change-points were detected by using the Pettitt test, Standard Normal Homogeneity test and Buishand test. Since an abrupt change in the series could cause a misleading outcome of the trend analysis, first we applied change-point detection. If at least one significant change appeared in the series, trend analysis was applied on each segment bounded by the change-points. Otherwise a trend analysis was applied to the whole series. Both long term and abrupt changes were applied also to overlapping time periods. Such analysis can provide better insight in the development of the changes which analyzing the whole series might not appear significant.

**Keywords:** Trend Analysis, Change-points, Hydrological Processes

**Acknowledgement.** This work was supported by Slovak Research and Development Agency under contract No. APVV-14-0013

## K-optimal designs for parameters of shifted Ornstein–Uhlenbeck processes

Sándor Baran<sup>1</sup>

<sup>1</sup>*Faculty of Informatics, University of Debrecen, Kassai street 26, H-4028 Debrecen, Hungary*

Continuous random processes are regularly applied to model temporal or phenomena in many different fields of science, and model fitting is usually done with the help of data obtained by observing the given process at various time points. In these practical applications sampling designs which are optimal in some sense are of great importance. We investigate the properties of the recently introduced K-optimal design [1] for temporal and spatial linear regression models driven by Ornstein–Uhlenbeck processes, and highlight the differences compared with the classical D-optimal sampling [2]. A simulation study displays the superiority of the K-optimal design for large parameter values of the driving random process [3].

**Keywords:** D-optimality, K-optimality, Optimal design, Ornstein–Uhlenbeck process

### References:

1. Ye, J. and Zhou, J. (2013) Minimizing the condition number to construct design points for polynomial regression models. *Siam. J. Optim.* 23, 666-686.
2. Zagoraiou, M. and Baldi Antognini, A. (2009) Optimal designs for parameter estimation of the Ornstein-Uhlenbeck process. *Appl. Stoch. Models Bus. Ind.* 25, 583-600.
3. Baran, S. (2017) K-optimal designs for parameters of shifted Ornstein-Uhlenbeck processes and sheets. *J. Stat. Plan. Inference* 186, 28-41.

## Data mining application issues in the taxpayers selection process

Mauro Barone<sup>1</sup>, Andrea Spingola<sup>1</sup>, and Stefano Pisani<sup>1</sup>

<sup>1</sup>*Italian Revenue Agency (Agenzia delle Entrate), Risk analysis and Tax compliance Research Unit, Via Cristoforo Colombo 426/D, Roma, Italy*

This paper provides a data analysis framework designed to build an effective learning scheme aimed at improving the Italian Revenue

Agency's ability to identify non-compliant taxpayers with special regard to sole proprietorship firms allowed to keep simplified registers.

Our procedure involves building two C4.5 decision trees, both trained and validated on a sample of 8,000 audited taxpayers, but predicting two different class values, based on two different predictive attribute sets.

That is, the first model is built in order to identify the most likely non-compliant taxpayers while the second one identifies the ones who more likely are not going to pay the additional due tax bill.

This twofold selection process target is requested in order to maximize the overall audit efficiency.

Once both models are in place, the taxpayer selection process will be held in such a way that businesses will only be audited if judged worthy by both models.

The methodology we suggest here will soon be validated on real cases: that is, a sample of taxpayers will be selected according to the classification criteria developed in this paper and will subsequently involved in some audit process.

**Keywords:** Data mining application, decision trees, tax fraud detection.

### **Greed and Fear: the Nature of Sentiment\***

**Giovanni Barone-Adesi<sup>1</sup>, Matteo M. Pisati<sup>2</sup>, Carlo Sala<sup>3</sup>**

*June 12, 2018*

Empirical indicators of sentiment are commonly employed in the economic literature while a precise understanding of what is sentiment is still missing. Exploring the links among the most popular proxies of sentiment, fear and uncertainty this paper aims to fill this gap. We show how fear and sentiment are specular in their predictive power in relation to the aggregate market and to cross-sectional returns. Finally, we document how sentiment and fear time cross-sectional returns: conditionally on a today's high (low) level of fear we observe a next month high (low) return per unit of risk. The opposite holds for sentiment.

**Keywords:** sentiment, uncertainty, fear, markets predictability, anomalies



## Revised survival analysis-based models in medical device innovation field

Andrea Bastianin<sup>1</sup>, Emanuela Raffinetti<sup>2</sup>

<sup>1</sup>*Department of Economics, Management and Statistics, University of Milano-Bicocca, Milan, Italy,* <sup>2</sup>*Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, Milano, Italy*

Scholars have shown that innovation and R&D affect both the business cycle and long-run economic growth (Basu et al. [1]; Comin and Gertle [2]). A statistical analysis of cross-country adoption of medical technology data, whose focus is on linear particle accelerators used as radiation treatment devices for patients with cancer, is presented. We exploit a unique database collecting information on some worldwide radiotherapy centres and concerning the exact year of medical device adoption, in order to compare the late-innovation functions of different groups of countries and to detect the basic economic, social and geographical features impacting on the early technological innovation opportunity. From a statistical point of view, a contribution to the study of technological medical innovations can be provided through the survival analysis-based models. Survival analysis resorts to both non-parametric and semi-parametric tools, such as the survival function (e.g. Kaplan and Meier [5]), which gives the probability of surviving beyond a certain event time  $t$ , and the Cox regression model (Cox [3]; Cox and Oakes [4]), which fulfills predictive purposes by detecting both the individual baseline hazard and that associated with the presence of specific factors impacting on the event occurrence. Typically, the event of interest takes a negative connotation since denoting a failure (e.g., length of time before a patient die after a disease). The fact that the survival and the cumulative distribution of a random variable are intertwined proves useful to interpret survival analysis results from an economic standpoint. Our proposal is to extend the survival analysis approach to the context of the innovative medical device adoption and its eventual diffusion within the worldwide countries, here representing the statistical units of interest. In such a perspective, a new perception of the main survival analysis tools is then provided. The event of interest is recognized in the initial adoption of a specific technology, becoming an indicator of medical technology innovation. On the contrary, the survival function is interpreted as an indicator of the delay in the technological innovation adoption since measuring the probability of introducing a novel medical device beyond a specific event time  $t$ . Given these features, the survival function is named late-innovation function. In

the same manner, also the Cox regression model is framed into an opposite scenario where the baseline hazard has no longer the meaning of risk but rather the meaning of early technological innovation opportunity, if no factors impacting on the initial technology adoption are taken into account. Analogously, the hazard function built on specific economic, social or geographical variables allows to detect their effects on the early technological innovation opportunity.

**Keywords:** medical device innovations, survival analysis, late-innovation function, early technological innovation opportunity

**References:**

1. S. Basu, J.G. Fernald, M.S. Kimball. Are Technology Improvements Contractionary ?. The American Economic Review, 96(5), 1418-1448, 2006.
2. D. Comin, M. Gertler. Medium-Term Business Cycles. American Economic Review, 96(3), 523-551, 2006.
3. D.R. Cox. Regression models and life tables. Journal of the Royal Statistical Society, Series B, 34, 187-220, 1972.
4. D.R. Cox, D. Oakes. Analysis of Survival Data. New York, Routledge, 1984.
5. E.L. Kaplan, P. Meier. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53, 457-481, 1958.

## Pre-emergence thermal and hydrothermal time model in crop

**Behnam Behtari**

*Department of Plant Ecophysiology, Faculty of Agriculture, University of Tabriz, East Azerbaijan, Iran*

Pre-emergence stage (germination to emergence) to determine the seedling emergence and its ecology plays an important role. So, calculating and verifying this critical stage of plant life is necessary for seedling emergence modeling. To calculate the length growth of seedlings in each day, first, the Hunt equation was used for calculating the relative growth rate of seedlings (RGRL).  $RGRL = (lnL_f - lnL_0) / t_{GE}$ . Where RGRL is relative growth rate,  $L_f$  the final length of shoot in mm,  $L_0$  the length at the onset of growth and  $t_{GE}$  the time spent from germination to emergence in h. RGRL unit was expressed as  $mm \cdot mm^{-1} \cdot h^{-1}$ . Assuming

that  $L_0=1$  mm just after germination, the final lengths of seedling were considered equal sowing depth and TGE value was calculated MSECE model, which is equal to  $t_E-t_G$ , the above equation can be written as:  $RGR_L = \ln\left(\frac{L}{L_0}\right) / (t - t_{GE})$ . The above equation with Gompertz, logistic and monomolecular models was fitted and the Gompertz model was selected to get the amount of length growth of seedlings from germination to emergence (EF 0.9 in most of the cases and RMSE<8.0). This equation was placed in thermal and hydrothermal time models so that the responses length growths from germination to emergence of these two important parameters are calibrated. In models of pre-emergence, sowing depth quantity was integrated with the soil temperature and water potential. However, under good agricultural practice, soil temperature and water potential are arguably the major influences on the timing and pattern of seed germination and seedling emergence in the field. Under more extreme conditions, seedbed factors such as increased soil impedance and reduced oxygen supply can have an overriding influence on seedling emergence. Here the situation is less complex as soil is prepared uniformly, and the whole population of seeds (usually non-dormant) is sown at the same time and to a uniform depth. Despite this, seedling emergence in crops is still variable and therefore directly influences both their yield and monetary value.

**Keywords:** Germination, Gompertz model, Logistic model, Monomolecular models

## Estimating the width of uniform distribution under measurement errors

**Mirta Benšić<sup>1</sup>, Safet Hamedović<sup>2</sup>, Kristian Sabo<sup>3</sup>**

<sup>1</sup>*Department of Mathematics, University of Osijek, Trg Ljudevita Gaja 6, Osijek, Croatia,*

<sup>2</sup>*Faculty of Metallurgy and Technology, University of Zenica, Travnička cesta 1, Zenica, Bosnia and Hercegovina*

<sup>3</sup>*Department of Mathematics, University of Osijek, Trg Ljudevita Gaja 6, Osijek, Croatia*

Recently, this group of authors has analyzed several statistical models related to the problem of estimating the uniform distribution length if the data were measured with an additive normal or Laplace error. It turned out that these models are of interest for applications in the problem of length

and area estimation for an object captured with error. Thus, an R package was developed which facilitates their usage. However, it appears in many images that heavier tailed error distribution gives better results than normal or Laplace. The main focus of the presentation is to describe the broad family of univariate distributions obtained by adding additive error to the uniformly distributed data with a special emphasis on estimating the width of the uniform support. The family has some useful properties regardless on the error distribution type as long as the error satisfies usual regularity conditions. We introduce the model, give some basic properties and conditions for maximum likelihood estimator to be asymptotically efficient and discuss robustness through simulated and real data.

**Keywords:** Uniform distribution, additive error, maximum likelihood estimator, robustness

## Nonparametric Regression Estimator for LTRC and Dependent Data

**Siham Bey<sup>1</sup>, Zohra Guessoum<sup>2</sup>, Abdelkader Tatachak<sup>3</sup>**

<sup>1</sup>Lab. MSTD, Faculty of Mathematics, University of Science and Technology Houari Boumediene, Algiers, Algeria. <sup>2</sup>Lab. MSTD, Faculty of Mathematics, BP 32, University of Science and Technology Houari Boumediene, Algiers, Algeria,

<sup>3</sup>Lab. MSTD, Faculty of Mathematics, BP 32, University of Science and Technology Houari Boumediene, Algiers, Algeria

Under left truncated and right censored (LTRC) model, a kernel estimator of the regression function is given for associated data. The local optimal bandwidth corresponding to the kernel is calculated by minimizing the mean squared error (MSE). We establish the strong uniform consistency on a compact set and give a rate of the almost sure convergence of the estimate.

**Keywords:** Truncated-censored data, Regression, Association, Kernel, Strong uniform consistency rate, Mean squared error, Optimal bandwidth

## Unimodality and Logconcavity of Density Functions of System Lifetimes

**Mariusz Bieniek<sup>1</sup>, Marco Burkschat<sup>2</sup>, Tomasz Rychlik<sup>3</sup>**

*<sup>1</sup>Institute of Mathematics, Maria Curie Skłodowska University, Pl. Marii, Poland, <sup>2</sup>Institute of Statistics, RWTH Aachen University, Aachen, Germany, <sup>3</sup>Institute of Mathematics, Polish Academy of Sciences, Warsaw, Poland*

We consider coherent systems composed of items with independent identically distributed absolutely continuous lifetimes. A special emphasis is laid on the case of components with uniform lifetimes. Then sufficient conditions on system signatures assuring logconcavity of the system lifetime density function are presented. They are weaker than logconcavity of the signatures. As a by-product, we obtain a positive answer to a classic analytic open problem if logconcavity of a function guarantees logconcavity of the respective Bernstein approximation operators. For the component lifetimes with general absolutely continuous distributions we show that logconcavity of the component lifetime density function and that of the system signature implies unimodality of system lifetime density function, but they are not sufficient for ensuring its logconcavity. Some assumptions on the component lifetime density function and system signature that provide the system lifetime logconcavity are described. Theoretical results are illustrated by examples.

**Keywords:** Coherent System, Signature, I.i.d. Component Lifetimes, Unimodal Density Function, Logconcave Density Function

## PageRank and Perturbed Markov chains

**Pitos Seleka Biganda<sup>1,2</sup>, Benard Abola<sup>2</sup>, Christopher Engström<sup>2</sup>, John Mango Magero<sup>3</sup>, Godwin Kakuba<sup>3</sup>, Sergei Silvestrov<sup>2</sup>**

*<sup>1</sup>Department of Mathematics, College of Natural and Applied Sciences, University of Dar es Salaam, Dar es Salaam, Tanzania, <sup>2</sup>Division of Applied Mathematics, The School of Education, Culture and Communication (UKK), Mälardalen University, Västerås, Sweden, <sup>3</sup>Department of Mathematics, School of Physical Sciences, Makerere University, Kampala, Uganda*

PageRank is a widely used hyperlink-based algorithm to estimate the relative importance of nodes in networks [1]. Since many real-world networks are large sparse networks, this makes efficient calculation of

PageRank complicated. Moreover, one needs to escape from dangling effects in some cases as well as slow convergence of the transition matrix. Primitivity adjustment with a damping (perturbation) parameter  $\varepsilon \in (0, \varepsilon_0]$  (for fixed  $\varepsilon_0 \approx 0.15$ ) is one of the essential procedures known to ensure convergence of the transition matrix [2]. If  $\varepsilon$  is large, the transition matrix loses information due to shift of information to teleportation matrix [3]. In this talk, we aim to formulate PageRank problem as the first and second order Markov chain perturbation problem. Using numerical experiments, we will compare convergence rates for the two problems for different values of  $\varepsilon$  on different graph structures and investigate the difference in ranks for the two problems.

**Keywords:** PageRank, Markov chains, Perturbation problem.

**References:**

- [1] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7), 107-117, 1998.
- [2] A. N. Langville and C. D. Meyer. *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2011.
- [3] D. Silvestrov and S. Silvestrov. *Nonlinearly Perturbed Semi-Markov Processes*. Springer, 2017.

## A Decomposition of Change in Disabled Life Expectancy at Retirement

Heather Booth<sup>1</sup> and Qi Cui<sup>2</sup>

*1 School of Demography, Australian National University, Coombs 9, Canberra, ACT 2601, Australia*

*2 School of Demography, Australian National University, Coombs 9, Canberra, ACT 2601, Australia*

As life expectancy advances, attention has turned to whether healthy life expectancy keeps pace. Do we live more or fewer years in a state of disability? The answer remains unclear and often differs between the sexes, producing the sex morbidity-mortality paradox that females have better survival but live more years in disability than males. Using an extended decomposition (Cui, Canudas-Romo and Booth), this paper examines the components of change in expected years lived in a disabled state at age 65. Based on age-specific rates for two events, disability and mortality, the three components capture the effects of differences between

disability and mortality in: change in rates, the life table age distributions of events, and expected remaining years in health and life by age. These three effects oppose and augment each other depending on relative change in disability and mortality rates. At low mortality levels, these effects tend to favour narrowing rather than widening of years in a disabled state. The method is applied to Australian data by sex from 1998 to 2012. Mortality rates are from the Human Mortality Database. Disability rates are from the Australian Bureau of Statistics cross-sectional Survey of Disability, Ageing and Caring, 1998-2012. Disabled life expectancy is calculated as the difference between life expectancy and disability-free life expectancy. For females over the 14-year period, life expectancy at 65 increased by 2.2 years including 1.4 years with disability (64%). For males, the overall increase was 3.0 years, also including 1.4 years with disability (46%). The decomposition reveals the trends in the three effects underlying change in disabled life expectancy, thereby enhancing understanding of this measure.

**Keywords:** Decomposition, Life Expectancy, Disability-Free Life Expectancy, Disabled Life Expectancy, Sex Morbidity-Mortality Paradox, Australia, Retirement.

## Efficiency of Brazilian Hospitals: assessment of Unified Health System (SUS)

Laura de Almeida Botega<sup>1</sup>, Mônica Viegas Andrade<sup>2</sup>, Gilvan Ramalho Guedes<sup>3</sup>

<sup>1</sup>*Departamento de Economia, Centro de Desenvolvimento e Planejamento Regional (Cedeplar), Universidade Federal de Minas Gerais (UFMG). Av. Pres. Antonio Carlos, Belo Horizonte MG, Brasil,* <sup>2</sup>*Departamento de Economia, Centro de Desenvolvimento e Planejamento Regional (Cedeplar), Universidade Federal de Minas Gerais (UFMG). Av. Pres. Antonio Carlos, 6627 Pampulha, Belo Horizonte MG, Brasil,* <sup>3</sup>*Departamento de Demografia, Centro de Desenvolvimento e Planejamento Regional (Cedeplar), Universidade Federal de Minas Gerais (UFMG). Av. Pres. Antonio Carlos, 6627 Pampulha, Belo Horizonte MG, Brasil*

This article assesses the efficiency for general hospitals that provides health services for the Brazilian Unified Health System (SUS). We used information on hospitalization and physical, human and financial

resources from the Hospital Information System (SIH/SUS) and the National Registry of Health Facilities (CNES). CNES is a national registry of mandatory completion that gathers monthly information of all health facilities, such as: capacity (for example: beds, equipments) and human resources (doctors, nurses, nursing assistants). The SIH/SUS is an administrative database that contains information regarding all hospitalizations financed by SUS including patients' characteristics (birth date, local of residence, gender) and medical procedures' characteristics (International Classification of Diseases – ICD, procedure classification, hospital code and localization). In total, 3,504 general hospitals were analyzed for the year 2015. We combined Data Envelopment Analysis (DEA) and Spatial Analysis to measure hospital inefficiency and to map its spatial pattern throughout the country. DEA is a linear programming approximation that estimates Decision Making Units (DMU) efficiency and compares to the best practice. One advantage of using DEA is that it allows estimating efficiency considering multiple inputs and multiple outputs. It also allows to decompose total efficiency in technical and scale efficiencies. We chose the input oriented model that demonstrates the extent to which inputs could be reduced in order to achieve efficiency. This analysis is more adequate since there is a budget constraint for public healthcare expenditure and therefore healthcare utilization could not be easily increased. We found a technical efficiency average score of 0.7134 and a scale efficiency average score of 0.6788. A considerable number of hospitals were operating at low levels of occupancy rate; the DEA slack analysis showed that many of these hospitals could increase production and reduce inputs to achieve higher efficiency standards. Hospital size and the type of healthcare provider (public, private and philanthropic) were associated with different patterns of efficiency. Most hospitals operated with increasing returns to scale, thus below the most productive scale level. We also found a positive association between hospitals' efficiency and municipalities' socio-economic indicators (population density, Gross Domestic Product, Municipality Human Development Index) as well as with patient's displacement distance. As higher is the average distance of patient's displacement to receive inpatient care the higher is the efficiency score. The spatial distribution of hospitals indicated that most of the inefficiencies were concentrated in areas away from the large urban centers and disproportionally represented by small hospitals. Medium and large hospitals presented, on average, higher levels of efficiency and occupancy rate, but most of them were still operating in a low capacity utilization rate. These findings suggest room to optimization, but



inequalities in access and the matching of demand and supply must be carefully considered in any attempt to reorganize the hospital system in Brazil.

**Keywords:** Hospital services, technical efficiency, scale efficiency, data envelopment analysis, spatial analysis

### Current multibloc methods. A comparative study in a unified framework

**Stéphanie Bougeard<sup>1</sup>, Ndèye Niang<sup>2</sup>, Thomas Verron<sup>3</sup>, Xavier Bry<sup>4</sup>**

<sup>1</sup>Department of Epidemiology, Anses, BP53, Technopole Saint Brieuc Armor, Ploufragan, France, <sup>2</sup>CEDRIC-CNAM, Paris, France, <sup>3</sup>SEITA, Paris, France, <sup>4</sup>Université de Montpellier, IMAG, Place Eugène Bataillon, Montpellier, France

The analysis of high-dimensional multiple datasets (i.e., when variables outnumber observations) consists in exploring and modelling the relationships between several blocks of variables measured on the same units. The blocks are connected by the user—with respect to some a priori information—and the connections are usually graphed on a path diagram. Many different methods have been developed to address this issue, such as PLS Path Modelling, regularized Generalized Structured Component Analysis (rGSCA), regularized Generalized Canonical Correlation Analysis (rGCCA), THEmatic Model Exploration (THEME) or, more recently, Path-ComDim, among others.

To help users understand the differences between these methods, we rewrote them in a common formal setting and compared them with respect to two key issues: (i) How do multiblock methods explore the block-relationships? (ii) How do multiblock methods separate information from noise? Underneath these two questions lie the notions of symmetrical or asymmetrical links, of balancing the goodness of fit with the component-strengths, of balancing the block-component-strengths in case of multicollinearity and of exploring the data space through multiple components. These methods have been applied to simulated data and to real data, to illustrate their differences and complementarities.

**Keywords:** component methods, structural equation, path-modelling, multiblock methods, dimension reduction

## A New Modified Scheme for Linear Shallow-Water Equations

**Aicha Boussaha**

*Department of Mathematics, National Superior School of Mines and Metallurgy,  
Annaba, Algeria*

We propose a modified scheme for simulating irregular wave trains (IWTs) propagation dispersive of tsunami with suitable initial and boundary conditions by applying the alternating direction implicit (ADI) method. The convergence, stability and consistency criteria of the scheme have been studied. We introduce a weakly dissipative terms into improved linear Boussinesq equations (ILBqs) that permits the mathematical tool to simulating a transoceanic propagation dispersive of tsunami in both ocean and laboratory experimental. The new numerical dispersion of the proposed model is manipulated to replace the physical dispersion of (ILBqs) by controlling dispersion-correction parameters. The new model developed in this study is applied to propagation of Heraklion tsunami scenario1 (HTS1) of the 365 AD earthquake. The resulting scheme is efficient and practical to implement. Furthermore, a comparison between the present results with another existing numerical method has been reported and we found that they are in a good agreement.

**Keywords:** Improved Linear Boussinesq equations; Numerical dispersion-correction parameter; ADI scheme; Dissipation effects; Tsunamis

### References

- [1] A. Boussaha, A. Laouar, A. Guerziz, Hossam S. Hassan, A new modified scheme for linear shallow-water equations with distant propagation of irregular wave trains tsunami dispersion type for inviscid and weakly viscous fluid, GJPAM(2015).
- [2] P. A. Madsen, R. Murray, O. R. Sørensen, A new form of the Boussinesq equations with improved linear dispersion characteristics, Coastal Eng., 15 (1991) 371-388.

## Redistricting Using Counties, Municipalities and the Convexity Ratio

**James R. Bozeman**

*American University of Malta*

In this work we propose a method of redistricting that disallows gerrymandering. Districts are formed by starting with entire counties (or similar regions), whose boundaries have unlikely been gerrymandered. If in order to meet the required population number a county must be subdivided then only entire municipalities are used, as their boundaries have probably not been gerrymandered. Furthermore, each district is created so that its convexity ratio is greater than or equal to .5, making the district nicely shaped, wherever possible. In the case of boundary districts this measure only occurs after adding the convex hull onto any unchangeable boundary lines. We also discuss the last step in the computer program calculating the convexity ratio for any given district.

**Keywords:** redistricting, gerrymandering, convexity ratio, boundary districts, counties and municipalities.

**References:** Bozeman JR, Davey M, Hutchins S, et al. Redistricting without gerrymandering, utilizing the convexity ratio, and other applications to business and industry. *Appl Stochastic Models Bus Ind.* 2018;1–17.

<https://doi.org/10.1002/asmb.2396>

## Modelling monthly birth and deaths using Seasonal Forecasting Methods as an input for population estimates

**Jorge Miguel Bravo<sup>1</sup>, Edviges Coelho<sup>2</sup>**

*<sup>1</sup>NOVA IMS, Universidade Nova de Lisboa, <sup>2</sup>Statistics Portugal, Portugal*

The Labour Force Survey (LFS) collects information on a sample population and, every calendar quarter, needs advanced data on estimates of resident population for each NUTS 3. In Portugal, the LFS quarter results are published around forty days after the end of the survey period. This calendar is incompatible with the current production of population estimates, since data on the three components – births, deaths and migration – are not yet available. As such, monthly forecasts of live

births, deaths and migration must be used. Empirical time series data for births and deaths by NUTS 3 in Portugal shows strong evidence of the presence of seasonality patterns, which mean that appropriate forecasting methods must be considered. In this paper we address the problem of forecasting monthly live births and deaths by NUTS 3 and sex and the distribution of the total predicted deaths by age. The purpose is to use seasonal forecasting methods in order to capture the seasonal behavior of the data. First, for each individual time series graphical analysis is used to analyze past behaviour of fertility and mortality. Second, three alternative methodologies are considered to model and forecast the number of births and deaths by NUTS 3 and sex: ARIMA models with a seasonal component, Holt-Winters exponential smoothing models, and state-space models. Multiple combinations of each of the three alternative types of models are used to fit births and deaths for each NUTS 3, and the best model is chosen using the BIC criterion. To evaluate the forecasting power of each model we use a back-testing procedure using various summary measures of the deviation between the observed values and the forecast point estimates. To assess the robustness of the empirical results to changes in the observation period, we conduct a sensitivity test on the forecasting power of each model considering a longer observation period and a more recent one. The methodology that provides the best forecasting performance for the majority of the NUTS 3 is adopted. Given the forecasted total number of monthly deaths for each NUTS 3, we use a cohort component approach to distribute deaths by individual age considering the most up-to-date death probabilities derived from complete life tables and a calibration procedure to redistribute the residual component.

Keywords: Births, Deaths, Forecast

## **New Dividends Strategies**

**Ekaterina Bulinskaya**

*Faculty of Mechanics and Mathematics, Lomonosov Moscow State University,  
Moscow, Russia*

New models describing the functioning of insurance companies have arisen in modern actuarial sciences during the 21st century (see, [1]). The models taking into account the dividends payment became very popular due to the fact that nowadays almost every insurance company is a stock-company. Hence, it is necessary to pay dividends to the shareholders. The

most well known strategy of dividends payment is the so-called barrier strategy. Many ramifications of this strategy are already developed (see, e.g., [2]). After the short historical survey we are going to propose some new strategies. They pertain to classical and dual continuous-time insurance models, as well as, discrete-time ones. It was established that in many situations such models are more appropriate than continuous-time ones. The typical case is dividends payment usually effectuated at the end of financial year. We also introduce the statistical estimates of model parameters and investigate their properties. Moreover, we analyze the models sensitivity to small parameters fluctuations and underlying distributions perturbations (see, [3], [4]).

**Keywords:** Dividends, Optimization, Stability, Simulation.

The research is partially supported by RFBR grant 17-01-00468.

#### References

- [1] Bulinskaya E. New Research Directions in Modern Actuarial Sciences. In: Modern Problems of Stochastic Analysis and Statistics, Springer, 2017, pp. 349-408.
- [2] Dassios A., Wu Sh. On Barrier Strategy Dividends with Parisian Implementation Delay for Classical Surplus Processes. Insurance: Mathematics and Economics, 2009, v. 45, pp. 195-202.
- [3] Bulinskaya E.V., Shigida B.I. Sensitivity Analysis of Some Applied Probability Models. Fundamental and Applied Mathematics. 2018, 8 (30), 19-34 (in Russian)
- [4] Rachev S.T., Klebanov L.B., Stoyanov S.V., Fabozzi F.J. The Methods of Distances in the Theory of Probability and Statistics. Springer, 2013.

## Goodness-of-Fit Testing for Point Processes in Survival Analysis

**Sami Umut Can<sup>1</sup>, Estate V. Khmaladze<sup>2</sup>, Roger J.A. Laeven<sup>3</sup>**

<sup>1,3</sup>*Department of Quantitative Economics, University of Amsterdam, Amsterdam, The Netherlands,* <sup>2</sup>*School of Mathematics & Statistics, Victoria University of Wellington, Wellington, New Zealand*

Suppose we are given an observed path from a temporal point process, and we would like to test whether a particular parametric model for the process' conditional intensity matches the observed path. We propose a novel approach to conducting such goodness-of-fit tests. The idea is to

consider the compensated point process, where the compensator is estimated parametrically, and to transform this process into a Poisson process, compensated by its own estimated compensator. Then it is sufficient to know the asymptotic behavior of the latter process in order to test the goodness-of-fit of the former, for a wide class of parametric intensity models. We demonstrate the applicability of our approach through Monte Carlo simulations of Aalen-type survival processes, with and without censoring.

**Keywords:** Point process, goodness-of-fit, martingale transform, survival analysis

## One bank problem in the federal funds market

**Elena Cristina Canepa, Traian A. Pirvu**

*McMaster University, 1280 Main S. W, Hamilton, ON, Canada*

**Description:** Systemic Risk in Banking and Finance.

The model of this paper gives a convenient strategy that a bank in the federal funds market can use in order to maximize its profit in a contemporaneous reserve requirement (CRR) regime. The reserve requirements are determined by the demand deposit process, modelled as a Brownian motion with drift. We propose a new model in which the cumulative funds purchases and sales are discounted at possible different rates. We formulate and solve the problem of finding the bank's optimal strategy. The model can be extended to involve the bank's asset size and we obtain that, under some conditions, the optimal upper barrier for fund sales is a linear function of the asset size. As a consequence, the bank net purchase amount is linear in the asset size.

**Keywords:** profit maximization, banking reserve requirements, bank's optimal strategy, Double Skorokhod Formula

## Calibration of two-factor stochastic volatility model

Betuel Canhanga<sup>1</sup>, Ying Ni<sup>2</sup>, Calisto Guambe<sup>1</sup>

<sup>1</sup>*Faculty of Sciences, Department of Mathematics and Computer Sciences, Eduardo Mondlane University, Mozambique,* <sup>2</sup>*Division of Applied Mathematics, Mälardalen University, Västerås, Sweden*

The two factor double mean-reverting stochastic volatility model proposed by Christoffersen et al. (2009) has the advantage of being more flexible in fitting the market implied volatility surfaces. Under the assumption of fast mean-reverting and slow mean-reverting respectively for the two stochastic volatility factors, Canhanga et al. (2016) has presented a calibration procedure for this model via a first-order asymptotic expansion approach. The calibration was done for real-market option data for one day. In the present study, we derive a second-order asymptotic expansion formula for calibration procedure, which is implemented for daily option data during one-year. We compare the calibration qualities between the first-order and second-order asymptotic expansion approaches. We compare also the asymptotic expansion approaches to more traditional calibration methods. Extensive numerical studies are performed to investigate the properties of the calibration procedures and for choosing an optimal calibration approach.

**Keywords:** Calibration, stochastic volatility model, multifactor stochastic volatility model, asymptotic expansion

### References:

- Christoffersen, C., Heston, S., and Jacobs, K. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science*, 55(12):1914–1932, 2009.
- Canhanga, B., Malyarenko, A., Murara, J.P., Ni, Y., Silvestrov, S. Numerical Studies on Asymptotics of European Option Under Multiscale Stochastic Volatility. *Methodology and Computing in Applied Probability*. vol 19 (4), December 2017

## The wide variety of regression models for lifetime data

**Chrys Caroni**

*Department of Mathematics, National Technical University of Athens, Zografou  
Campus, Athens, Greece*

The dependence of lifetimes on covariates can be modelled in a regression format using various techniques that have been introduced in the literature on lifetime data analysis (survival analysis, reliability modeling). In some fields, notably biostatistics, the most familiar approach is through the proportional hazards assumption, especially as implemented in Cox's semi-parametric version. Also well known, but less widely used, is the proportional odds model. There are other "proportional" models in the literature, notably proportional reversed hazards and proportional mean residual life. Another model based on hazards is the additive hazards model. In the field of reliability, the accelerated failure time model is the dominant methodology. Another approach, possessing the conceptual advantage of being based on a representation of the underlying process leading to the event (such as death or failure), is threshold regression in which (in the Inverse Gaussian case) two parameters of the underlying distribution are modeled. Burke & Mackenzie have also introduced a multi-parameter regression.

In this paper, we present and discuss these models and emphasize the differences between them and the circumstances in which each is useful. Various theoretical relations are well known, such as the fact that only the Weibull distribution can be characterized either as a proportional hazards model or as an accelerated failure time model, and only the log-logistic distribution as either proportional odds or accelerated failure time. We present further results on the relationship between the different approaches.

**Keywords:** Lifetime data; Regression models; proportional hazards; proportional odds; accelerated failure time; threshold regression



## Quantization of Transformed Lévy Measures

**Mark Anthony Caruana**

*Department of Statistics and Operations Research, Faculty of Science,  
University of Malta*

In this paper we find an optimal approximation of the measure associated to a transformed version of Lévy-Khintchine canonical representation via a convex combination of a finite number  $P$  of Dirac masses. The quality of such an approximation is measured in terms of the Monge-Kantorovich or the Wasserstein metric. In essence, this procedure is equivalent to the quantization of measures. This procedure requires prior knowledge of the functional form of the measure. However, since this is in general not known, then we shall have to estimate it. It will be shown that the objective function used to estimate the position of the Dirac masses and their associated weights (or masses) can be expressed as a stochastic program. The properties of the estimator provided are discussed. Also, a number of simulations for different types of Lévy processes are performed and the results are discussed.

Keynotes: Quantization of Measures, Lévy-Khintchine Canonical representation, Stochastic Programming, Wasserstein Metric, Monge-Kantorovich Metric

## Modeling of mortality in elderly by trachea, bronchus and lung cancer diseases in the Northeast of Brazil

**João Batista Carvalho<sup>1</sup>, Neir Antunes Paes<sup>2</sup>**

*<sup>1</sup>Department of Statistics, Federal University of Campina Grande, R. Aprígio Veloso, 882 - Universitário, Campina Grande - PB, 58429-900, Brazil, <sup>2</sup>Health and Decision Modeling Postgraduate Course. Federal University of Paraíba, Campus I - Lot. Cidade Universitária, João Pessoa - PB, Brazil*

In 2015, for the elderly (60 and over), trachea, bronchus and lung (TBL) cancer was the second leading cause of cancer death in Brazil (14%) and in the Northeast of the country (12.5%). With a population of 56 million inhabitants, the Northeastern Brazil is considered one of the less developed regions of the country and the Latin America. The main goal was to identify sociodemographic and socioeconomic determinants of mortality in older people by TBL cancer applying the Structural Equations Modeling (SEM). It was adopted a cross-sectional ecological study using micro-data information from 2010 census and projected population for

2015 linked to register data on cause-specific mortality by TBL by sex to the 188 micro regions of the Northeast. The following steps were performed: 1) The death data distribution by age were corrected; 2) Age-standardized corrected mortality rates were computed; 3) The SEM was applied. The outcome mortality rates due to TBL cancer were observed directly, and a set of indicators regarding to health, education, income and environmental conditions were used indirectly as latent variable. The SEM proved to be highly sensitive with significance in the measurement model for some relevant latent variables that can subsidize political planning. Rates were higher in microregions with lower percentages of illiterate elderly and in poverty, lower dependency ratio and higher percentage of elderly people living in households with running water. The levels were higher in elderly men and increased with the age. In view of the high rates observed, and an aging process and life expectancy rising in the Northeast, it is expected that these levels should increase even more in the near future.

**Keywords:** Lung Diseases, Cancer, Mortality in the Elderly, Mortality by Causes

### On a family of risk measures based on largest claims

Antonia Castaño<sup>1</sup>, Gema Pigueras<sup>2</sup>, Miguel Ángel Sordo<sup>3</sup>

<sup>1,2,3</sup> *Department of Statistics and Operation Research, University of Cádiz, Spain,*

In an insurance framework, we introduce a family of premium principles in terms of the expected average risk of the largest claims in a set of independent and identically distributed claims. Each premium principle of this family can be represented by mixtures of tail value-at-risks, with beta mixing distributions. From this representation, we obtain a convergence result that connects the tail value-at-risk with the largest claims of a portfolio. A characterization of the excess-wealth order in terms of this family of premiums is provided. As a consequence, we obtain a sufficient condition for ordering the net premiums of two collective risks under the ECOMOR reinsurance treaty.

**Keywords:** risk measure, premium principle, order statistics, excess-wealth order, reinsurance

## A general piecewise multi-state survival model for the study on the progression of breast cancer

Juan Eloy Ruiz-Castro<sup>1</sup> and Mariangela Zenga<sup>2</sup>

<sup>1</sup>*Department of Statistics and Operational Research and IEMath-GR. University of Granada. Faculty of Science. Campus Fuentenueva s/n. 18071, Spain.*

<sup>2</sup>*Department of Statistics and Quantitative Methods, University of Milano-Bicocca. Via Bicocca degli Arcimboldi, 8, 20126, Milano, Italy*

Multi-state models are considered in the survival field for modeling illnesses which evolve through several stages over time and they can be developed by applying several techniques, such as non-parametric, semi-parametric and stochastic processes, Markov processes in particular. When the development of an illness is being analyzed, its progression is tracked in a periodic form. In this work, we show a non-homogeneous piecewise Markov process in discrete time for a three-state model for relapse and survival times for breast cancer patients who have undergone mastectomy.

**Keywords:** Multi-state model, stochastic model, discrete time, breast cancer.

## Performance estimation of a wind farm with a copula dependence structure

Laura Casula<sup>1</sup>, Guglielmo D'Amico<sup>2</sup>, and Giovanni Masala<sup>3</sup> and Filippo Petroni<sup>4</sup> and Robert Adam Sobolewski<sup>5</sup>

<sup>1</sup>*Dipartimento di Scienze Economiche e Aziendali, Università degli studi di Cagliari, 09123 Cagliari, Italy.*

<sup>2</sup>*Dipartimento di Farmacia, Università 'G. D'Annunzio' di Chieti-Pescara, 66013 Chieti, Italy.*

<sup>3</sup>*Dipartimento di Scienze Economiche e Aziendali, Università degli studi di Cagliari, 09123 Cagliari, Italy.*

<sup>4</sup>*Dipartimento di Management, Università Politecnica delle Marche, 60121 Ancona, Italy.*

<sup>5</sup>*Bialystok University of Technology, Bialystok, Poland.*

The production of energy by a wind power plant is closely connected to the intensity of the wind at the site where the plant is located. To correctly model the energy produced, it is therefore necessary to consider a model that can faithfully reproduce the intensity of the wind. In this regard, we use a power transformation of the wind speed data where the optimal exponent is determined such that the transformed values are close to a

Gaussian distribution (see Sim et al. [1] for further details). Then, we apply a SARIMA process to the transformed data in order to cope with seasonality and residual autocorrelation. Next, energy production can be deduced from wind speed by considering the power curve that characterizes the wind turbine. This model permits indeed to replicate the statistical behavior of the energy production of a single wind turbine.

Then, we must replicate the statistical features of the electricity price series. The model foresees a deterministic and a stochastic component. The deterministic component copes with the seasonal behavior of the series and it is well represented by a combination of trigonometric functions. The stochastic component is modeled with an Ornstein-Uhlenbeck process (mean-reverting) with the inclusion of jumps. The jump process is able to capture the price peaks which characterizes the electricity prices (see also Weron [2] for a survey of electricity price modeling).

The final aim is to estimate performance and losses related to the production of electricity deriving from unexpected fluctuations in energy production. To this end, we observe that the price of electricity is correlated, on an hourly basis, to the intensity of the wind (and therefore, consequently, to energy production). To take this dependency into account, we apply a copula function.

A numerical application of this model is performed on real data coming from a wind farm located in Italy (Sardinia). We compare then real data with synthetic data obtained by implementing the proposed model. We then estimate the indicators of interest such as the Loss of load expectation and the consequent economic performance and loss coming from unexpected fluctuations of power production

**Keywords:** Wind energy, Copula dependence structure, Loss of load expectation, Economic performance indicators.

#### **Acknowledgements**

Authors Laura Casula and Giovanni Masala are supported by Fondazione di Sardegna and Regione Autonoma della Sardegna (L.R. 7/2007 annualità 2016).

#### **References**

1. Sim, S.-K., Maass, P. & Lind, P.G. (2019). Wind Speed Modeling by Nested ARIMA Processes, *Energies* 12, 69; doi:10.3390/en12010069.
2. Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future, *International Journal of forecasting* 30, pp. 1030-1081.

## Bayesian analysis for the system lifetimes under Frank copulas of Weibull component lifetimes

Ping Shing Chan, Yee Lam Mo

*Department of Mathematics and Statistics, The Hong Seng University of Hong Kong, Shatin, Hong Kong*

In this paper, we consider a coherent system of  $n$  Weibull lifetime components where the joint distribution of these  $n$  components is represented by the Frank copulas. Given a random sample of  $m$  system lifetimes, we analyze the data via Bayesian inference by assuming the prior distribution of the parameters to be known. The posterior distribution of the unknown parameters is obtained by the Metropolis Hastings within Gibbs algorithm. A numerical simulations will be presented to illustrate the proposed method.

**Keywords:** Bayesian Computation; Gibbs Sampler; Frank Copulas; Reliability; MCMC algorithm

## Psychometric validation of constructs defined by ordinal-valued items

Anastasia Charalampi<sup>1</sup>, Catherine Michalopoulou<sup>2</sup>, Clive Richardson<sup>3</sup>

<sup>1</sup>*Postdoctoral Fellow, Department of Social Policy, Panteion University of Social and Political Sciences,*

<sup>2</sup>*Professor of Statistics, Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece*

<sup>3</sup>*Emeritus Professor of Applied Statistics, Department of Economic and Regional Development, Panteion University of Social and Political Sciences, Athens, Greece*

Determining the structure and assessing the psychometric properties of constructs (scales) before their use is a prerequisite of scaling theory. This involves splitting randomly a sample of adequate size into two halves and first performing Exploratory factor analysis (EFA) on one half-sample in order to assess the construct validity of the scale. Secondly, the structure suggested by EFA is validated by carrying out Confirmatory factor analysis (CFA) on the second half. Based on the full sample, the psychometric properties of the resulting scales or subscales are assessed. The appropriate methods of analysis depend on the level of measurement of

the items defining the scale. In this paper, we carry out the investigation and assessment of the 2012 European Social Survey (ESS) short eight-item version of the Center for Epidemiologic Studies Depression (CES-D) scale for Italy and Spain when items are considered as ordinal. In both countries, EFA performed on the first half-samples resulted in a one-factor solution comprised of six items, as the factor with the remaining two items was considered as poorly defined. CFA performed on the second half-samples and the full samples resulted in adequate model fit for both countries.

Although we did not confirm the unidimensionality of the short eight-item version of the CES-D scale, a single subscale that was both reliable and valid was identified. Further research is necessary in every country and both rounds of the ESS that included the depression measurement (2006 and 2012) in order to establish subscales suitable for use in analyses.

**Keywords:** Depression; Exploratory factor analysis; Confirmatory factor analysis; Reliability; Construct validity; European Social Survey

## **Calculation of Analogs of Lyapunov Indicators for Different Types of EEG Time Series Using Artificial Neural Networks**

**German Chernykh, Ludmila Dmitrieva, Yuri Kuperin**

*St. Petersburg State University, Russia*

The method of constructing numerical characteristics of time series, which are analogs of Lyapunov indicators for dynamic systems is developed in the present paper. The method is applied to a comparative analysis of EEG time series for subjects in states of meditation and background. Lyapunov indicators are related to the rate of divergence of close phase trajectories. Consequently, the calculation of the Lyapunov indicators requires the availability of information on several scenarios of the evolution of a dynamic system with slightly different initial conditions. In order to obtain such information for time series, in this work multilayered artificial neural networks of direct propagation (MLP) are used, which are trained on the EEG time series under consideration. The training of neural networks and the calculation of Lyapunov indicators is preceded by the procedure of embedding EEG time series into the lag space. The reconstructed attractor in the lag space is topologically equivalent to an attractor in real phase space for dynamical system generated the time

series under consideration. Thus, for the EEG time series, the spectrum of analogs of the Lyapunov indicators for the trajectories on the reconstructed attractor is calculated. It is shown that under compliance with the main principle of neural networks learning, which consists in minimizing the learning error without neural network retraining, the values of Lyapunov indicators obtained by the method described above, within the statistical error do not depend on the parameters of neural networks used to generate alternative trajectories. A comparative analysis of Lyapunov indicators allows us to talk about obtaining statistically significant results, on the basis of which one can judge whether the subject is in a state of meditation or in a background state. More detailed neurophysiologic explanations of the obtained results will be presented in the talk.

**Keywords:** Lyapunov Indicators, Artificial Neural Networks, Meditation, Reconstructed Attractor

## **Long Term Care Insurance in Singapore: Assessing the Shield Index**

**Ngee Choon Chia**

*Department of Economics, Faculty of Arts and Social Sciences, National University of Singapore*

Long-term care (LTC) insurance in Singapore, or CareShield Life, is currently the only social safety net in place to combat healthcare expenditure arising from LTC needs. It explores the adequacy of LTC insurance in reducing LTC cost with actuarial modelling and simulations. Using data from the Ministry of Health and other survey studies, we construct a multi-state model based on the Markov process and input transition probabilities for different health states to perform actuarial calculations. Monte-Carlo simulations are subsequently conducted to assess the adequacy of Eldershield, where we found that only 13% of LTC cost can be covered with ElderShield payouts. Finally, we investigate the impact of two policy changes to enhance the coverage of ElderShield and relax the trigger factor. These changes will significantly improve the comprehensiveness of the LTC insurance.

## Real Estate Market Analysis with Time-frequency Decomposition

Siu Kai Choy, Tsz Fung Stanley Zel

*The Hang Seng University of Hong Kong, Shatin, Hong Kong*

The real estate market is one of the most complex and important element in the financial market. Abrupt changes in housing prices usually cause significant impact to the economy, especially inflation. These changes are possibly driven by government policies, economic and financial events. Detecting such changes accurately and studying their characteristics would help investors to understand the relationship among government policies, economic events and their financial implications. While existing abrupt change detection methodologies have been applied successfully to various financial time series, they may not perform well in abrupt change detection for the signals with the presence of noise. To remedy this shortcoming so as to improve the detection accuracy, a combined approach that integrates the Hilbert-Huang transform with a newly developed multi-level rate-of-change detector for abrupt change detection is proposed. The proposed method aims to simultaneously perform signal decomposition and reconstruction, and capture abrupt changes in the noisy signal. Comparing to the existing methodologies, the proposed method achieves a remarkable higher accuracy in terms of detection of abrupt changes.

**Keywords:** Real estate market analysis, abrupt change detection, time-frequency decomposition



## Variability and the latent ageing process in life histories: Applications on cohort studies

**M.D. Christodoulou, J.A. Brettschneider, D. Steinsaltz,**

*Department of Statistics, University of Oxford*

*24-29 St Giles' Oxford, UK*

How do we age as individuals? What are the hidden trajectories each of us follow as we grow old? How much of that process is genetic and how much is it the reflection of our experiences and chosen lifestyles? To answer complex questions such as these we need to combine complicated datasets with novel statistical methods. Longitudinal studies, such as the British Birth Cohorts, are rich, carefully curated datasets with great potential. Collected through the decades, these large datasets have been used to provide us with snapshots of the lives we lead. In a topic as complex as ageing however, a snapshot is simply not enough. We live varied lives, have different experiences and different genetic endowments, and as such we age differently. To study our life trajectories we need to exploit the available datasets to their full extent. The primary limiting factor to such exploitation is the lack of appropriate statistical tools. Our developed tools build on the intuition that beneath all the ups and downs of an individual life there is a clock that measures the individual rate of ageing. Following on from our pilot study on the fertility ageing patterns of the participants of the 1958 British Birth Cohort, we test our methodology on other longitudinal studies such as the Wisconsin Longitudinal Study. This helps us assess our methods under a variety of conditions and examine which of the patterns we observed in the 1958 Cohort are present in other groups and which are not. We extend this further by carefully assessing the impact of metrics employed to summarise socioeconomic status on our findings and use our developed pipeline for polygenic score calculation to add a genetic layer to our interpretation.

**Keywords:** ageing, cohort studies, 1958 British birth cohort, polygenic scores, socioeconomic status, fertility, reproductive ageing, latent processes

## Polya-Aeppli Geometric process

Stefanka Chukova<sup>1</sup>, Leda D. Minkova<sup>2</sup>

<sup>1</sup>*School of Mathematics, Statistics Victoria University of Wellington, NZ,*

<sup>2</sup>*Faculty of Mathics and Informatics, Sofia University "St. Kl.Ohridski", Bulgaria*

In this talk we introduce a new point process called Polya-Aeppli geometric process (PAGP), with underlying exponential distribution. We provide the system of differential equations for the distribution of the number of events of the PAGP up to time  $t$  and discuss some of its properties. The new process is an extension of the well known Polya-Aeppli process, as well as the standard geometric process with underlying exponential distribution.

**Keywords:** Polya-Aeppli distribution, Polya-Aeppli process, geometric process

**Acknowledgement.** The second author was partially supported by Grant DN12/11/20.dec.2017 of the Ministry of Education and Science of Bulgaria.

## Latent Class Analysis in the psychological research on attachment styles and the transition to parenthood

Franca Crippa<sup>1</sup>, Mariangela Zenga<sup>2</sup>, Rossella Shoshanna Procaccia<sup>3</sup>, Lucia Leonilde Carli<sup>1</sup>

<sup>1</sup>*Department of Psychology, University of Milano-Bicocca, Milano, Italy,*

<sup>2</sup>*Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milano, Italy,*

<sup>3</sup>*University Ecampus, Novedrate-Como, Italy*

Latent Class Analysis (LCA) allows the identification of unmeasured class membership among subjects using either categorical or continuous observed variables, or both. It can be considered a mixture modelling since the probability distribution of the overall population is a mixture of the distributions of the subpopulations. Referring to LCA as a mixture model is useful to underline that it is a person-centred, or person-oriented, approach, i.e., it describes similarities and differences among individuals, rather than among variables. <sup>1</sup>Department of Psychology, University of Milano-Bicocca, Italy. In this study we consider the application of LCA to data from a psychological research carried out in Lombardy, a Northern Italian region. The research focussed on the transition to parenthood, enshrining every sort of generative choice, not to have children too, in relation to the attachment styles in different phases of the life cycle. The R environment is used.

**Keywords:** Latent Class Analysis, Mixture Modelling, Transition To Parenthood, Attachment

## Alcohol consumption in selected European countries

Jana Vrabcová<sup>1</sup>, Kornélia Svačinová<sup>2</sup>, Markéta Pechholdová<sup>3</sup>

<sup>1</sup> Department of Statistics and Probability, University of Economics, Prague, Czech Republic

<sup>2</sup> Department of Demography, University of Economics, Prague, Czech Republic

<sup>3</sup> Department of Demography, University of Economics, Prague, Czech Republic

**Background:** Alcohol is the most widespread psychoactive substance with negative direct and indirect effects on both the consumer and his broad environment. Compared to the rest of the world, alcohol consumption in Europe is generally high, but large disparities persist in consumption levels and patterns.

**Aims:** This paper assesses the alcohol consumption in Europe from the perspective of recorded per capita alcohol consumption and self-reported drinking patterns. Our aim is to combine information from multiple data sources to reveal similarities and differences in drinking, including gender, educational and income disparities.

**Data and methods:** WHO and OECD data was used to measure alcohol consumption. Drinking patterns, including heavy episodic drinking, were analysed based on the European Health Interview Survey from 2014, which has been recently made accessible online. Descriptive analysis, frequencies and correlations were used as analytical tools.

**Results:** Alcohol consumption levels have homogenized across Europe and decreased in many countries to rank between 6-12 litres of pure alcohol per capita in 2016. Several patterns of regular and episodic drinking were identified with respect to gender, educational and income disparities. In most countries however, men drink more than women, more educated and better situated drink more and riskier than poor with low education.

**Conclusions:** Two trends arise from our analysis: compared to the past, alcohol consumption stabilized at much lower levels, and there are no signs of a future dramatic increase. However, the spread of alcohol in well-situated populations and excessive female drinking in countries with high gender equity suggests that alcohol reducing policies should also aim at higher social groups.

**Keywords:** Alcohol, Mortality, Morbidity, European countries, Czech Republic

## **A proportional hazard model under bivariate censoring and truncation**

**Hongsheng Dai, Chao Huang, Miriam J. Johnson, Marialuisa Restaino**

*University of Salerno, Fisciano, Italia*

The bivariate survival data are usually subject to incomplete information due to censoring and truncation. Most existing works focused on estimating the bivariate survival function when only one component is censored or truncated and the other is fully observed. Only recently bivariate survival function estimation under the assumption that both components are censored and truncated has received considerable attention. Moreover, the most common approaches to model covariates effect on survival time are the Cox PH and AFT models, that have been well studied for the univariate censored data. Not much has been done for the bivariate survival data when truncation is present. The paper aims at estimating the regression coefficients in the bivariate proportional hazards model, when both components are censored and truncated. In particular, truncation is considered as covariate in the regression model, in order to evaluate its effect on the hazard estimation. A simulation study and an application to real data are conducted to investigate the performance of the estimators.

**Keywords:** Bivariate survival data, truncation, censoring, Cox model

## Simultaneous Threshold Interaction Modeling Approach for Paired Comparisons Rankings

**Antonio D'Ambrosio<sup>1</sup>, Alessio Baldassarre<sup>2</sup> and Claudio Conversano<sup>2</sup>**

<sup>1</sup> *Department of Economics and Statistics, University of Naples Federico II, Napoli - Italy*

<sup>2</sup> *Department of Business and Economics, University of Cagliari, Cagliari, Italy*

Rankings and paired comparisons rankings are ubiquitous in data analysis. Recently, in the literature there has been a growing interest in modeling rank data, especially in trying to explain and/or predict preferences explicitly stated from a sample of judges starting from a set of covariates over a set of alternatives. Both parametric and non-parametric tools have been introduced in order to deal with preference rankings or paired comparison rankings as response variable.

In this work we introduce a model dealing with the identification of threshold interaction effects in paired comparisons rankings response data, which integrates recursive partitioning and generalized linear and non-linear models for preference rankings.

**Keywords:** STIMA, Paired comparisons, Preference rankings, trunk modeling.

## A Probabilistic Model of Wind Farm Power Generation via Copulas and Indexed Semi-Markov Models

**Guglielmo D'Amico<sup>1</sup>, Giovanni Masala<sup>2</sup>, and Filippo Petroni<sup>3</sup> and Robert Adam Sobolewski<sup>4</sup>**

<sup>1</sup> *Dipartimento di Farmacia, Università 'G. D'Annunzio' di Chieti-Pescara, 66013 Chieti, Italy.*

<sup>2</sup> *Dipartimento di Scienze Economiche e Aziendali, Università degli studi di Cagliari, 09123 Cagliari, Italy.*

<sup>3</sup> *Dipartimento di Management, Università Politecnica delle Marche, 60121 Ancona, Italy.*

<sup>4</sup> *Bialystok University of Technology, Bialystok, Poland.*

We face the problem of modelling the wind power production of a wind farm composed of a given number of wind turbines. Due to geomorphological structure of the land, to the shear effect induced by the blades and to other physical factors, it is well known that the stochastic

properties of the total produced energy cannot be obtained simply considering the produced power of a single turbine with the total number of turbines. Therefore, we aim to develop a complete model that is able to correctly reproduce and forecast the power production of the whole wind farm. At this purpose, we represent the stochastic production of energy of each turbine using an indexed semi-Markov chain (ISMC). This choice is suitable to reproduce the statistical properties of power production of a single wind turbine.

The modelling of the whole wind farm is performed by introducing a dependence structure through a copula function. This allows us to get to the heart of the issue disposing of a multivariate process that describes the wind power of each single wind turbine as well as the existing interdependencies among the wind turbines.

**Keywords:** semi-Markov, wind energy, risk measure, Monte Carlo simulation, copula,

### **Rocof of higher order for multistate systems in continuous time**

**Guglielmo D'Amico<sup>1</sup> and Filippo Petroni<sup>2</sup>**

<sup>1</sup> *Dipartimento di Farmacia, Università "G. d'Annunzio" di Chieti-Pescara, Chieti, Italy*

<sup>2</sup> *Dipartimento di Management, Università Politecnica delle Marche, Ancona, Italy*

In this paper we study the rate of occurrence of failures (ROCOF). The ROCOF has been considered by Shi (1985) for finite Markov processes (MP). Lam (1997) determined a formula for the ROCOF of a MP or of a higher-dimensional MP admitting the possibility to work with a denumerable state space. A further extension to semi-Markov processes (SMP) was advanced by Ouhbi and Limnios (2002). The ROCOF gives information whether there are a lot of failures or only a few within a time interval. In the study of failures of a system, it is also interesting the study of the relative positioning of pairs of failures and more in general of tuples of failures. Consequently, an extension of the ROCOF, called ROCOF of higher order, was calculated for MP in D'Amico (2015). Here we consider SMP and a mixed probability distribution for the initial law of the system taking into account the possible random starting from any state of the system with any duration. Furthermore, we determine a formula for the ROCOF of higher order for SMP and we recover as particular cases the

results contained in the above quoted papers. An application demonstrates how to implement applications.

**Keywords:** Rocof, Semi-Markov Processes, reliability.

### **A copula based Markov Reward approach to the credit spread in European Union**

**Guglielmo D'Amico<sup>1</sup>, Filippo Petroni<sup>2</sup>, Philippe Regnault<sup>3</sup>, Stefania Scocchera<sup>1</sup>, Lorian Storch<sup>1</sup>**

<sup>1</sup>*Department of Pharmacy University G. D'Annunzio Chieti - Pescara, Chieti Italy,*

<sup>2</sup>*Department of Management University Politecnica delle Marche, Ancona, Italy,*

<sup>3</sup>*Laboratory of Mathematics University of Reims Champagne- Ardenne, UFR SEN Moulin de la Housse Reims, France*

In the present work, we propose a methodology based on piecewise homogeneous Markov chain for credit ratings and a multivariate model of the credit spreads to evaluate the financial risk in European Union (EU). Two main aspects are considered: how the financial risk is distributed among the European countries and how large is the value of the total risk. The first aspect is evaluated by means of the expected value of a dynamic entropy measure. The second one is solved by computing the evolution of the total credit spread over time. Moreover, the covariance between countries' total spread allows understand any contagions in the European Union. The methodology is applied to real data of 24 European countries for the three major rating agencies: Moody's, Standard & Poor's and Fitch. Obtained results suggest that both the financial risk inequality and value the total risk increases over time at a different rate depending on the rating agency and that the dependence structure is characterized by a strong correlation between most of European countries.

**Keywords:** Sovereign credit rating, Markov process. Dynamic measure of inequality, Copula, Change-point

## The Generalized Calculation of Pure Premium for Non-Life Insurance by Generalized Non-Homogeneous Markov Reward Processes

**Guglielmo D'Amico<sup>1</sup>, Jacques Janssen<sup>2</sup>, Filippo Petroni<sup>3</sup>, Raimondo Manca<sup>4</sup> and Ernesto Volpe Di Prignano<sup>4</sup>**

<sup>1</sup>*Università G. D'Annunzio, 66013 Chieti, Italy*

<sup>2</sup>*Université Libre de Bruxelles, Belgium*

<sup>3</sup>*Università Politecnica delle Marche, 60121 Ancona, Italy*

<sup>4</sup>*Università di Roma "La Sapienza", Roma, Italy*

The paper describes the calculation, in a simple and precise way, of the aggregate claim amount and the claim number using generalized Markov reward models in a non-homogeneous time environment. The evolution equations of the generalized non-homogeneous Markov reward processes in a discounted environment is the tool that will be used for the calculation of the aggregate claim amount and in a non-discounted case for the calculation of the mean claim number. The paper can be considered as a generalization of the paper D'Amico et al. (2017) that proposed the introduction of the age as main time variable.

The main new results presented in this paper are:

- The simultaneous introduction inside the model of the calendar time as second time variable, the introduction of two time variables gives the opportunity of:
- A more precise usage of the non-homogeneous setting, as will be explained inside the paper,
- The possibility of following, in a better way, the evolution equation of a non-life insurance contract,
- The evolution of technology,

The evolution of inflation and consequently costs and earnings of each cohort of an Insurance Company, where in a cohort there are all the insureds with the same age, sex and driving experiences. In the last section, will be presented a non-life Insurance application. The database that will be used for the calculation of the aggregate claim amount and the mean claim number will be done using a database with more than 2000000 records in which each record reports the history of driving behavior of each insured. The result will show the importance of the age of the insured people and the possibility to consider simultaneously the calendar time in the calculation of the actuarial quantities.

**Keywords:** aggregate claim amount process; claim number process; Markov chains, age, calendar time, reward process, non-homogeneity.



## Dynamic inequality: a Python tool to compute the Theil inequality within a stochastic setting

Guglielmo D'Amico<sup>1</sup>, Stefania Scocchera<sup>2</sup>, Lorian Storch<sup>1</sup>

<sup>1</sup> *Department of Pharmacy University G. D'Annunzio Chieti - Pescara, via dei Vestini 31 Chieti Italy*

<sup>2</sup> *Department of Management University Politecnica delle Marche, Piazza Martelli 8 Ancona, Italy*

The present work aims at presenting a software for the evaluation of dynamic measures of inequality on the attribute's distribution among a set of  $N$  individuals. The amount of this attribute depends on a discriminatory criterion, according to whom the individual belongs to alternative groups, i.e. the meta-communities. Two main approaches have been implemented. Firstly, the Markov reward approach allows to model attribute evolution as a reward process driven by the underlying meta-community, which evolves according to discrete-time homogeneous Markov chain. Secondly, a Copula-based Markov reward approach generalizes the first one by including the multivariate modeling of the attributes. The software is then completed by the implementation of the change-point detection algorithm which enables the meta-community process to be piecewise.

After the description of the methodology, this work gives detailed description of the Command Line Interface (CLI) along with the Graphical user Interface (GUI) built in order to make the software ready and easy to use for all researchers. Finally, several applications are presented to show the potential usefulness of this tool. The software is freely available at the present URL: <https://github.com/lstorchi/markovctheil>

**Keywords:** Python, Markov process. Dynamic measure of inequality, Copula, Change-point, Attribute, Meta-community.

## On the number of observations in random regions determined by records

Anna Dembińska<sup>1</sup>, Masoumeh Akbari<sup>2</sup>, Jafar Ahmadi<sup>3</sup>

<sup>1</sup>Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland <sup>2</sup>Department of Statistics, University of Mazandaran, Babolsar, <sup>3</sup>Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

Near-record observations are the ones that occur between successive record times and within a fixed distance of the current record value. In this talk, we will generalize the concept of near-record observations to the notion of observations that fall into a random region determined by a given record and a Borel set. We will describe the exact distribution of the number of such observations and establish limiting properties of this number. In addition, we will give some asymptotic results for sums of such observations. Numbers of observations falling into random regions determined by records are interesting not only from the theoretical point of view as natural extensions of numbers of near-record observations. They can find applications in different fields. During the talk we will indicate some applications of the presented theoretical results in hydrology, meteorology, insurance and record theory. In particular, we will use the new results to derive exact and asymptotic properties of inter-record times and of numbers of repetitions of records.

**Keywords:** Records, Near-record observations, Inter-record times, Repetitions of record, Limit theorems

## Expected lifetimes of coherent systems with DNID components

Anna Dembińska<sup>1</sup>, Agnieszka Goroncy<sup>2</sup>

<sup>1</sup>Warsaw University of Technology, Warsaw, Poland, <sup>2</sup>Nicolaus Copernicus University in Toruń, Poland

We consider a set of possibly dependent and not necessarily identically distributed (DNID, for short) discrete random variables. For such a setting we compute moments of respective order statistics and present applications of this result to reliability theory. We introduce a method which allows for establishing expectation of lifetimes of coherent systems consisting of possibly dependent and nonhomogeneous components. We

focus on the case of systems with multivariate geometrically distributed components' lifetimes.

**Keywords:** order statistics, coherent system, system's lifetime, dependent and not necessarily identically distributed random variables, discrete distribution

### **Residual lifetime of k-out-of-n: G systems with a single cold standby unit**

**Anna Dembińska<sup>1</sup>, Nikolay I. Nikolov<sup>2</sup>, Eugenia Stoimenova<sup>2</sup>**

*<sup>1</sup>Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland, <sup>2</sup>Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. B, Sofia, Bulgaria*

In this study, we consider a k-out-of-n: G system with a single cold standby unit. This system consists of n components and it functions if at least k of its components work. When the system fails for the first time, i.e. at the time when the (n-k+1)th failure occurs, the standby unit immediately replaces one of the failed components and is put in operation. We assume that the system operates in cycles or is monitored at discrete time and that the component lifetimes have discrete joint distribution. In order to describe the aging behavior of the system we consider the expectations of the residual lifetime of the system, the residual lifetime at the system level and the residual lifetime given that the system is still working. Since the calculation of these characteristics requires to find a sum of infinite series, we provide a procedure to approximate them with an error not greater than a fixed value  $d > 0$ . The special cases when the component lifetimes have geometric, negative binomial and discrete Weibull distributions are studied in details.

**Keywords:** Reliability, k-out-of-n: G system, Mean residual lifetime, Discrete lifetime distribution, Approximation procedure

## Co-clustering for time series based on a dynamic mixed approach for clustering variables

Christian Derquenne

*Electricité de France, Research and Development, Palaiseau, France*

In many applications, the search for structures in the data is essential to explore and visualize, to understand the formation of phenomena, to identify subsets of information in order to apply adapted treatments (modeling, forecasting, ...). The search for patterns in the data is at the heart of this problem; it uses unsupervised classification methods. These are usually applied to either variables or individuals. However more and more applications require the simultaneous construction of classes of variables and individual (co-clustering): microarray in bioinformatique, video content recognition, users and movies in recommender systems, customers and days of electricity consumption curves,...and the number of methods proposed is growing. We will place ourselves in the context of the co-clustering of daily electricity consumption curves of customers to explain the new approach that we propose. The variables correspond to 48-hour days and the individuals are the customers. The objective is to build daily profiles of customers in order to offer them adequate rates for their consumption behavior. Our method contains three steps. The first ones is to classify the variables (the days) knowing the information from the customers, while the second step groups the individuals (the customers) knowing the days. The first step uses a dynamic mixed approach for clustering variables (Derquenne, 2016, 2017) for the days, then a Multiple Correspondence Analysis (MCA) or dissimilarity matrix (DM) to obtain a global clustering of days for overall customers. The second step uses a classical criteria of clustering (Ward, complete linkage,...) for the customers, then a MCA or a DM is applied to obtain a clustering of customers for overall days. The third step implements an approach to reconcile information from the previous two steps to obtain of crossed-cluster customers and days. Our approach is applied to simulated data and actual electricity consumption data.

**Keywords:** Co-clustering, unsupervised learning, dissimilarity index, Multiple Correspondence Analysis, time series

## Estimating gross margins for agricultural production in the EU: approaches based on the equivariance of quantile regression

**Dominique Desbois**

*UMR Economie Publique, INRA-AgroParisTech, Université Paris-Saclay, Paris, France*

This communication introduces the estimation by product of the agricultural production gross margins using the equivariance property of the quantile regression with an application to member countries of the European Union. After recalling the conceptual framework of the estimation of agricultural production costs, the first part, presents the equivariance property of quantile regression and its consequences for the estimation of gross margins on the basis of specific costs for agricultural production. The second part documents the data collection used by this estimation procedure and distributional characteristics of specific costs for productions of twelve Member States of the European Union. According to a comparative analysis between the member states, the third part presents the econometric results of some major products using factor analysis and hierarchic clustering based on the related estimation intervals. The last section discusses the relevance of these methodological approaches using various criteria.

**Keywords:** input-output model, agricultural production cost, micro-economics, quantile regression, equivariance, factor analysis, hierarchic clustering, interval estimates

### References:

- Afonso F., Diday E., Toque C. (2018). Data science par analyse des données symboliques. Une nouvelle façon d'analyser les données classiques, complexes et massives à partir des classes - Applications avec Syr et R, Technip editions, Paris, 442 p.
- Desbois D. (2015) Estimation des coûts de production agricoles : approches économétriques. ABIES-AgroParisTech PhD Thesis, directed by J.C. Bureau & Y. Surry, 249 p.
- Desbois D., Butault JP., Surry Y. (2017). Distribution des coûts spécifiques de production dans l'agriculture de l'Union européenne : une approche reposant sur la méthode de régression quantile, *Economie rurale*, n° 361, pp. 3-22.

## Pricing of Longevity derivatives and cost of capital

Pierre Devolder<sup>1</sup>, Fadoua Zeddouk<sup>2</sup>

<sup>1</sup>*Institute of Statistic, Biostatistic and Actuarial Science (ISBA) Université Catholique de Louvain (UCL), Louvain la Neuve, Belgium,* <sup>2</sup>*PHD student UCL University, Belgium*

The significant improvement in longevity in most developed countries increases the annuity providers' exposure to longevity risk. In order to hedge this risk, new longevity derivatives have been proposed (longevity bonds, q-forwards, survivor swaps, options...). Although academic researchers, policy makers and practitioners have talked about it for years, the longevity-linked derivatives available in the financial market are still limited, in particular due to the pricing difficulty. In this paper, we compare different existing pricing methods and propose a Cost of Capital approach based on economic capital arguments, following Levantesi and Menzietti (2017) framework. Our method is designed to be more consistent with Solvency 2 requirement i.e. the Solvency Capital Required should cover with 99.5% probability the unexpected losses on a one-year time horizon. The price of longevity risk is determined for a S-forward and a S-swap but can be used to price other longevity-linked securities. To describe mortality we use affine stochastic processes, and we show that mean reverting models with a time-dependent level are more appropriate to describe the death intensity of individuals. In particular, the Hull & White and CIR extended models are used to represent the evolution of mortality over time. We use data for Belgian population to derive prices for the proposed longevity linked securities based on the different methods.

**Keywords:** Stochastic longevity risk, S-forward, S-swap, Cost of Capital, mean reverting models

## A Flexible Regression Model for Compositional Data

Agnese M. Di Brisco<sup>1\*</sup>, Sonia Migliorati<sup>1</sup>

<sup>1</sup>*Department of Economics, Management and Statistics - University of Milano-Bicocca, Milan, Italy*

Compositional data on the simplex are defined as vectors with strictly positive elements subject to a unit-sum constraint. The aim of this contribution is to propose a regression model for multivariate continuous variables with bounded support by taking into consideration the flexible

Dirichlet (FD) distribution which can be interpreted as a special mixture of Dirichlet distributions. The FD distribution is an extension of the Dirichlet one, which is contained as an inner point, which enables a greater variety of density shapes in terms of tail behavior, asymmetry and multimodality. A convenient parameterization of the FD is provided which is variation independent and facilitates the interpretation of the mean vector of each mixture component as a piecewise increasing linear function of the overall mean vector. A multivariate logit strategy is adopted to regress the vector of means, which is itself constrained to sum-up to 1, onto a vector of covariates. Intensive simulation studies are performed to evaluate the fit of the proposed regression model particularly in comparison with the Dirichlet regression model. Inferential issues are dealt with by a (Bayesian) Hamiltonian Monte Carlo algorithm.

**Keywords:** mixture model, proportions, multivariate regression

## Reverse Mortgages: Risks and Opportunities

**Emilia Di Lorenzo, Gabriella Piscopo, Marile Sibillo, Roberto Tizzano**

*University of Naples Federico II, Via Cintia, Naples, Italy*

Reverse mortgages (RM) provide an attractive way to increase retirement incomes and to face the needs of health care for elderly people. The RM market is exposed to a number of risks: (1) longevity risk, as retirees' life expectancy increases, (2) interest rate risk, especially in the low-rate post-crisis period, (3) property market risk, in the last stage of the current business cycle. We measure the overall risk for an insurer with a book of RM contracts. We also evaluate the optimal demand for RM in a retiree's investment portfolio, taking into account diversification with respect to other asset classes. Numerical results are shown and suggestions to further developments of the market are offered.

**Keywords:** Reverse Mortgage, Financial risk, Demographic risk

### References:

Alai D. H., Chen H., Cho D., Hanewald K, Sherris M.: Developing Equity Release Markets: Risk Analysis for Reverse Mortgage and Home Reversion. *North American Actuarial Journal*, 18:1, (2014) pp. 217-241.  
Chen H., Cox S. H., Wang S. S.: Is the Home Equity Conversion Program in the United States sustainable? Evidence from Pricing Mortgage Insurance Premiums and Non Recourse Provisions Using the Conditional

Escher Transform. Insurance: Mathematics & Economics, 46 (2010) pp. 371-384.

de la Fuente Merencio I., Navarro E., Serna G.: Estimating the No-Negative-Equity Guarantee in Reverse Mortgages: International Sensitivity Analysis. in New Methods in Fixed Income Modeling. Springer International Publishing DOI: 10.1007/978-3-31995285-7\_13 (2018)

Fornero E., Rossi M., Virzì Bramati M. C.: Explaining why, right or wrong, (Italian) households do not like reverse mortgage. Journal of Pension Economics & Finance, Vol. 15, Issue 2 (2016) pp.180-202.

Institute and Faculty of Actuaries: Lifetime Mortgage. A good and appropriate investment for life companies with annuity liabilities?" May 2014

## **Introduction of reserves in self adjusting steering the parameters of a pay-as-you-go pension plan**

**Keivan Diakite<sup>1</sup>, Abderrahim Oulidi<sup>2</sup>, Pierre Devolder<sup>3</sup>**

*<sup>1</sup>International University of Rabat (UIR), Retirement & Pension Chair, Parc Technopolis Rabat-Shore, Sala el Jadida, Morocco, <sup>2</sup> International University of Rabat (UIR), Retirement & Pension Chair, Parc Technopolis Rabat-Shore, Rocade Rabat-Salé - 11 100 Sala el Jadida, Morocco, <sup>3</sup>Université catholique de Louvain (UCL), Institute of Statistic, biostatistic and Actuarial Science (ISBA), Louvain la Neuve, Belgium*

The demographic trend of pension funds in Morocco (increased longevity combined with a drop in birth rates) and the situation of the labour market (a large share of the informal sector in employment) are a major challenge for the future of pay-as-you-go pension schemes in Morocco.

The mandatory Moroccan pension system operates in provisioned distribution and is financed by defined benefits. In the past, the surplus situation of the various plans has allowed them to accumulate significant financial reserves (22% of the GDP in 2016). In order to adapt the structural challenges related to shortfalls in defined benefit management, several parametric reforms have been carried out, each time in order to postpone the date of exhaustion of reserves. Projections show that future parametric reforms will be unsustainable in terms of contribution rates or career extensions. Thus it appears that a structural overhaul of the Moroccan pension system is more than necessary. This study focuses on



the transformation of the current a pay-as-you-go system in defined benefits into a pension points managed system and the introduction of a rule of automatic piloting of the different parameters of the regime over time. : The Musgrave rule, and then to stochastically model the introduction of reserves as a variable controlling the parameters of the regime. To do this, we present the theoretical framework of the Musgrave rule in the management and control of the regime in point as well as the effect of the introduction of reserves. In a second part we will present the architecture of the Moroccan pension system as a whole before focusing on the largest public pension fund by presenting its characteristics, parametric reforms and their impact. We will then simulate the transformation of the fund into a pension-managed plan by applying the Musgrave rule and the introduction of reserves. Finally we compare the current system with the new simulated system by measuring the impact of this transformation on the level of benefits and contributions through contribution rates and replacement rates.

**Keywords:** Benefits, Moroccan retirement system, Musgrave rule, Reserves, Pension

## **The Distribution of the Inverse Cube Root Transformation of Error Component of the Multiplicative Time Series Model**

**Dike Awa<sup>1</sup>, Chikezie David<sup>2</sup>, Otuonye Eric<sup>2</sup>**

*<sup>1</sup>Department of Maths/Statistics, Akanu Ibiam Federal Polytechnic, Unwana, P.M.B. 1007, Afikpo, Ebonyi State, Nigeria, <sup>2</sup>Department of Statistics, Faculty of Biological/Physical Sciences, Abia State University, P.M.B. 2000, Uturu, Nigeria*

This work examines the inverse cube root transformation of error component  $e_t^* \left[ = \frac{1}{\sqrt[3]{e_t}} \right]$  of multiplicative time series model. The probability density function (pdf) of the inverse cube root transformation of the multiplicative time series model was established, Further the  $f(y)$  was proved as a proper pdf since  $\int_0^\infty f(y)dy = 1$ . . The result was validated using the  $n^{th}$  root transformation. The result conform to the general rule. The moments (mean and variance) of the inverse cube root transformation were equally established using the general rule.

**Keywords:** Power Transformations, Probability Density Function Error Component, Mean Variance, Multiplicative Time Series

**Computation of the optimal policy for a two-compartment single vehicle routing problem with simultaneous pickups and deliveries, stochastic continuous demands and predefined customer order**

**Theodosios D. Dimitrakos<sup>1</sup>, Constantinos C. Karamatsoukis<sup>2</sup>,  
Epaminondas G. Kyriakidis<sup>3</sup>**

*<sup>1</sup>Department of Mathematics, University of the Aegean, Karlovassi 83200, Samos, Greece, <sup>2</sup>Department of Military Sciences, Hellenic Military Academy, Vari 16673, Attica, Greece, <sup>3</sup>Department of Statistics, Athens University of Economics and Business, Patission 76, 10434, Athens, Greece*

We develop and analyze a mathematical model for a specific stochastic vehicle routing problem. A vehicle with finite capacity starts its route from a depot and visits  $N$  customers in order to deliver to them new products and pick up from them old quantities of the same product. It is assumed that the vehicle has two compartments. We name these compartments, Compartment 1 and Compartment 2. The new products are stored in Compartment 1 and the old products are stored in Compartment 2. The vehicle must deliver and pick up products according to a predefined customer order. For each customer the quantity of the products that is delivered and the quantity of the products that is collected are continuous random variables with known distributions. The actual demands for new products and for old products are revealed upon the arrival to customer's site and cannot exceed the capacity of Compartment 1 and of Compartment 2, respectively. The vehicle is allowed to return to the depot to restock with new products and to unload the old products. It is possible to find the optimal routing strategy by implementing a suitable dynamic programming algorithm. Numerical examples for our problem are also presented.

**Keywords:** Vehicle routing problem, Dynamic programming, Logistics, Pickup and delivery, Compartments, Stochastic continuous demands

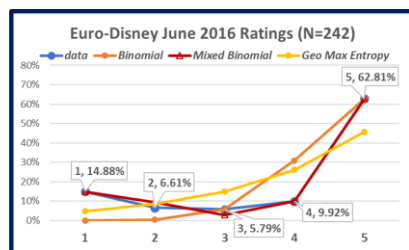
## Entropy Analytics of Euro-Disney Facebook Five Star Rating Dataset

Yiannis Dimotikalis

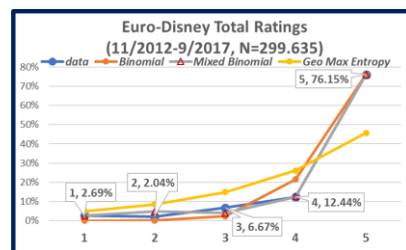
*Department of Management Science and Technology, Hellenic Mediterranean University, Agios Nikolaos, Crete, Greece*

This paper analyzes a dataset of about 300.000 five-star ratings by visitors of Euro-Disney in Paris in the time period 11/2012-9/2017. The distribution of those ratings to five-star scale analyzed monthly and yearly applying max entropy distribution principle. Geometric max entropy distribution, Binomial and mixed Binomial distributions applied to specific parts of dataset. At that time period several phenomena affecting the ratings of visitors-tourist happen, mainly the terrorist attacks in European cities and specifically in Paris and 2016 UEFA European Championship. The outcomes of those expected and unexpected events in visitors' rating is measured and examined in time evolution of ratings. Observed that unexpected negative events affect tourism and the opinion of tourists expressed by their rating (see fig. 1 and fig. 2 for the differences). The proposed mixed Binomial distribution is capable to represent almost perfectly the distribution of ratings.

**Keywords:** Max Entropy Distribution, Geometric Distribution, Binomial Distribution, Facebook ratings, five-star rating, Euro-Disney, unexpected events.



**Fig. 1:** Distribution of Euro-Disney Visitors Ratings, June 2016



**Fig. 2:** Distribution of Euro-Disney Visitors Ratings, all period 11/2012-9/2017

## **Application of Modified Local Holder Exponents Method to Study the Multichannel EEG in States of Meditation and Background**

**Ludmila Dmitrieva, Yuri Kuperin, German Chernykh, Daria Kleeva**  
*St. Petersburg State University, Russia*

A new method of studying multichannel EEG series, namely the method of Modified Local Holder Exponents (MLHE), is used in the present work. The algorithm of MLHE has been elaborated by some authors of the present research in order to analyze the regularity of time series around any given point, but it always has been applied to the study of financial time series. The non stationarity of the EEG records has been overcome by splitting the entire EEG time series for each channel into stationary sections by moving windows. The first aim of the work was to calculate MLHE for each stationary sections of EEG series for subjects in states of meditation and background. The second aim was to find statistically significant difference in the quantitative characteristics of MLHE corresponding to EEG in states of meditation and background. It is known that the greater the values of the MLHE time series, the more smooth is the series under consideration. The main parameters that were studied in the MLHE time series were the mean, median and mode. It turned out that practically for all channels we have statistically significant differences in these parameters. Namely, the values of mean, median and mode are greater for MLHE time series in meditative state than that in background state. This means that EEG recordings in a meditative state are smoother than in a background state. From a neurophysiologic point of view it means that the level of physiological noise in large neural ensembles significantly decreases for in a state of meditation. Also this means that synchronization of large neural ensembles is greater in meditative state than in background state.

**Keywords:** Modified Local Holder Exponents, Multichannel EEG series, Meditation

## Stochastic Modeling of Affinity Based Cognitive Networks

Denise Duarte<sup>1</sup>, Gilvan Guedes<sup>2</sup>, Gilvan Guedes<sup>3</sup>, Rodrigo Ribeiro<sup>4</sup>

<sup>1</sup>Department of Statistics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, <sup>2</sup>Demography Department, Av. Antonio Carlos 6627, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, <sup>3</sup>Department of Statistics, Av. Antonio Carlos 6627, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil,

<sup>4</sup>Department of Mathematics, Universidad Catolica de Chile, Santiago, Chile

In this work, we propose a stochastic modeling of cognitive networks based on elements of graph theory. We start supposing a finite dictionary,  $D$ , with support given by an induction term (as used in the Free Word Association Technique), which is common for all subjects in a population. From this dictionary, we can build a list,  $D^k$ , with all sequences of  $k$  words, for a given  $k$ . Then we measure how similar two sequences in  $D^k$  are through an affinity function,  $\alpha$ , between them. For example, we may consider the affinity function that returns one if two sequences share at least one word and zero otherwise. Any fixed affinity function defines a connection between pairs of sequences in  $D^k$ . In this way, we can view the set  $R = (D^k, E_\alpha)$  as a deterministic graph, where  $E_\alpha$  is the set of edges generated by the affinity function  $\alpha$ . We call this graph as the *relational map* of the dictionary  $D$ , for sequences of  $k$  words, induced by the affinity function  $\alpha$ .

Let us now consider a finite sample  $X_1, \dots, X_n$  from a random variable assuming values in the relational map  $R$  where each  $X_i$  can choose a node of  $R$  according to a probability measure  $\mu_i$ . The realization of a sample is, therefore, a subgraph of  $R$  with  $n$  nodes. We study the properties of this random subgraph investigating how a particular choice of the probability distribution  $\mu_i$  influences the general behavior of the sample graph.

**Keywords:** Vocabulary, Affinity, Graph, Probability Model, Cognitive Network

## **Comparative study of two different SW-RPA approaches to calculate the excess entropy of some liquid metals**

**N.E. Dubinin**

*1)Ural Federal University*

*2)Institute of Metallurgy of the Ural Division of the Russian Academy of Sciences, Ekaterinburg, Russia*

The square-well (SW) model for the pair intermolecular interaction is widely used to describe different types of matter in liquid and amorphous states. On the other side, to investigate different model fluids, the methods of the thermodynamic perturbation theory are intensively used. Among them, the random phase approximation (RPA) occupies a worthy place. In particular, in majority of works where the SW model is applied to metal state, the RPA is used. Here, we compare two different SW-RPA expressions obtained for the excess entropy in works [1, 2]. Calculations are fulfilled for seven pure liquid metals. It is found that these two formulas give results near each other and a good agreement with experiment. From the forms of the formulas under consideration, it allows to conclude that the used values of the SW parameters lead to a small difference between the square-well and hard-sphere structure factors. The work was supported by Act 211 Government of the Russian Federation (contract № 02.A03.21.0006) and by the Fundamental Research Program of UB RAS (project 18-10-3-28).

**Keywords:** Entropy; square-well model; thermodynamic perturbation theory; random phase approximation; liquid metal

### **References:**

1. M. Itoh, J. Phys. C: Solid State Phys. 20 (1987), 2483.
2. N. E. Dubinin, A.A. Yuryev, N.A. Vatolin, J. Non-Equilibr. Thermodyn. 35 (2010), 289.

## **Correcting death rates at advanced old age: review of models**

**Dalkhat M. Ediev**

*North Caucasian State Academy (Russia), Lomonosov Moscow State University (Russia), International Institute for Applied Systems Analysis (Austria).*

We address the problem of improving the quality of demographic data on mortality and expectation of life at old age. The leading cause of

inadequate quality of old-age mortality statistics is inaccurate recording of age of respondents at censuses, surveys and registers. Resolving this problem was a subject matter of scholar dispute in the literature (Horiuchi and Coale 1982; Coale 1985; Mitra 1985; Mitra 1984). Analysis of the alternative models based on more complete and comparable data (Ediev 2018) revealed that approached by Horiuchi-Coale and Mitra are, in fact, in a good consistence with each other and lead to substantial improvement of the accuracy of traditional estimates of expectation of life at old age. This, in turn, enables reducing, by several times, errors of extrapolative models of age-specific mortality (Ediev 2017). Unfortunately, the mentioned models and are not universally applicable and their fit is reduced as life expectancy increases worldwide. Those models, in particular, were assuming population stability that is not relevant to the current state in developed countries including Russia. We aim at developing more realistic mathematical models and methods for cases where the models available now are not applicable. Adjusting the Mitra model, combining known models, developing regression and behavioral models, in particular, seem to offer much improved estimates of old-age mortality and life expectancy.

The research leading to these results has received funding from the Russian Foundation for Basic Research under Grant 18-01-00289 “Mathematical models and methods of correcting the distortions of the age structure and mortality rates of elderly population”.

### **Estimating differences between two models based on different input data and environmental factors**

**Christopher Engström**

*Division of Applied Mathematics, The School of Education, Culture and Communication (UKK), Mälardalen University, Västerås, Sweden*

When comparing the performance between two propulsion systems in practice it is often impossible or too costly to test them under the same conditions. For large sea vessels, even a very small reduction in fuel consumption of 2-3% corresponds to a large financial gain, but to test a new system you generally only have the option of comparing performance during normal shipping hours. Generally, this means the two systems have to be compared during different sea conditions, velocities, shipload and various other environmental factors.

In this paper we will consider the “best case” scenario where you have two similar models which perfectly describes the underlying dynamics in the presence of no measurement errors. We will look at how accurately it is possible to compare the difference between the two models when model parameters are estimated under different input data with added measurement errors of different distributions. Of particular interest is the case with added bias on parts of the data representing long term unknown environmental factors such as differences in shipload.

**Keywords:** Monte-Carlo method, Least-Squares, Estimation

## Cumulative Density Function from Contaminated Noise

Ben Jrada M Es-salih<sup>1</sup>, Djaballah Khadidja<sup>2</sup>

<sup>1,2</sup>*Faculty of mathematics department of probability and statistics, University of Sciences and Technology Houari Boumediene Algiers Algeria*

Let  $\{X_i\}_1^{+\infty}$  be a strictly stationary stochastic process for positively associated random variables. We consider the deconvolving estimation of the Cumulative Density Function (CDF)  $F(x)$  of  $X$ . But the variable  $X$  is not available to observe directly, instead of  $X$  we observe  $Y=X+e$  where  $e$  is the measurement error with a known distribution. Given  $n$  observations of  $Y$ , we consider  $F_n(x) = \int_{-\infty}^x f_n(t) dt$ , as an estimator of  $F(x)$  where  $f_n(t)$  is the deconvolution density estimator of  $f(t)$ . The underlying process  $\{X_i\}_1^{+\infty}$  is assumed to satisfy certain mixing conditions, The asymptotic properties of the CDF estimator depends heavily on the smoothness of the noise distributions, The deconvolution kernel estimator of the CDF is therefore dependent on the choice of the two parameters :deconvolution kernel  $\omega$ , and the bandwidth  $h_n$ , we will see that the bandwidth parameter plays crucial role in order to get good asymptotic properties. The asymptotic rate of convergence corresponding to the bias of the CDF estimator are given. We noted that the bias of this estimator is the same to that based on direct observations ( $Y=X$ ) ie when no observation noise is present, and it's converges to zero in the booth cases. In order to trait the precise asymptotic rate and constant of the variance, We consider the tail of the characteristic function of the noise process  $e$  decays algebraically at infinity(ordinary smooth case), and Noted that the presence of contaminating noise reduce the mean-square convergence



rate of  $F_n(x)$  by a factor that depends on the rate of decay of the tail characteristic function of the noise process  $\{e_i\}_{i=1}^{+\infty}$

**Keywords:** Deconvolving estimate, Quadratic-mean Convergence and Rates, Positively Associated

## Identification of thyroid cancer risk factors incidence in urban and rural areas, Pakistan

Asif Faiza<sup>1</sup>, Muhammad Noor-ul-Amin<sup>2</sup>

<sup>1,2</sup>Department of Statistics

COMSATS University Islamabad, Lahore

As like the other countries, the risk of thyroid carcinoma is significantly increasing over the last few decades in Pakistan. This study aspires to know the cause and effect of this disease in urban and rural areas by investigation of different risk factors. For this purpose, incidence data was collected from Institute of Nuclear Medicine & Oncology Lahore and Sheikh Zayed Hospital, Lahore. This study consists of 88 rural and 232 urban patients and the possible risk factors of thyroid cancer investigated via questionnaire. The logistic regression is used as a statistical tool and the results are computed on the behalf of odd ratios. The result shows that 48 rural and 112 urban cases are suffered from thyroid cancer. In rural areas two factors use of iodine diet and oxidative stress are seen to be significant with odd ratios 1.642 and 1.796 while in the urban areas seven factors residential Area, oxidative Stress, too much consumption of meat & fast food, too much use of crucifer vegetables, excess use of fats and sea food are seen to be significant with odd ratios 0.760, 2.121, 1.294, 1.187, 1.618, 1.632 and 0.892, respectively. It is observed that the oxidative stress is the common factor in urban and rural areas.

**Keywords:** Thyroid carcinoma, Oxidative stress, Iodine diet, Crucifer vegetables, Odds Ratio

## **Changes of the Tehran City Floating Population Based on the 2006 and 2011 Census Data**

**Reza Sotoudeh Farkosh<sup>1</sup>, Ashraf Mashhadi Heidar<sup>2</sup>**

*<sup>1</sup>M.A in Demography, Statistical Centre of Iran, Tehran, Iran,*

*<sup>2</sup>B.S in statistics, Statistical Centre of Iran, Tehran, Iran*

The purpose of this study is to review the changes of the Tehran floating population based on the 2006 and 2011 census data. In this study which is an applied and a review one in consideration with its goal and execution (documentary-library), 2006 and 2011 census data for Tehran city were obtained from the Statistical Centre of Iran (SCI). The results showed that the floating population in Tehran is about 7.2 percent, while this proportion for the urban population of Tehran province except Tehran city and for the urban population of the total country was about 2 and 9 percent, respectively. This difference indicates that the population of urban areas around Tehran city, is affected by the floating population less than Tehran city. The vast majority of the floating population to Tehran city (99 percent) have an urban origin. Floating population of Tehran has increased by 3.1 percent between the years 2006 and 2011. This situation shows the increasing attractions of Tehran city in terms of employment and education for the demographic areas of the country and in particular, the regions around Tehran city. The motive for about 84 percent of the floating population who move to Tehran is finding job and the rest of them is to study, while this proportion is about 69 and 53 percent in the urban employed population of Tehran province and the total country employed population, respectively. The majority of the floating population who move to Tehran for the purpose of finding job, has an urban origin and men include a 6-fold share of the urban floating population in Tehran as compared with women. As the age grows, there is a decrease in the floating employed population.

Besides, about 16 percent of the floating population move to Tehran with the aim of study, while this proportion is about 35 percent in the urban student floating population to Tehran province. Since the results of this study indicate that the floating population of Tehran city is increasing, it could be mentioned that ignoring this phenomenon may lead to inconsistency in urban planning. Therefore, it would be necessary to provide appropriate strategies for Tehran city as a metropolis in order to decentralize and control floating population and tackling the existing problems.

**Keywords:** Floating Population, Tehran City, Employment, Job Status, Educational Status

## **Predicting The Customers Trend in Digital Firms Case Study in Iran**

**Saeed Fayyaz**

*Statistician on ICT statistics*

*Statistical Center of Iran (SCI), Dr. Fatemi, Tehran, Iran*

The digital economy in its broadest sense has transformed the way societies work and communicate. Remarkably, firms today are able to capitalize on digital tools to revolutionize production processes and/or to sell goods and services to markets that were previously out of scope, providing significant benefits to consumers through greater choice and cheaper prices. Online purchasing has been increasing dramatically in Iran and change the enterprises' figures. In this paper, an ARIMA model for predicting the customers trend in a digital-base firm has been broadened. In the second step, based on location data of customer destinations, a predicting model fitted to help decision makers for madding geographical decisions. This model shows the distribution of customers' destinations and it can be beneficial for future relational models between the type of products purchasing by customers and the places that they are located. Due to the competitive environment in digital marketing, this issue has turned to critical matter for such these firms. In order to reach actual and practical result, a big and successful firm which has offered both products and services online, was selected as a case study. As far as inner-joins relation between the decisions and customers, this study will be helpful in similar companies to improve both the customers' volume and the products.

**Keywords:** Digital Purchasing, Time Series Model, Location-Based Data, Decision Making.

## Is Taylor's power law true for random networks?

István Fazekas, Csaba Noszály, and Noémi Uzonyi

University of Debrecen, Faculty of Informatics, 4028 Debrecen, Kassai street 26,  
Hungary

Taylor's power law states that the variance function is quadratic. It is observed for population densities of species in ecology. Taylor's power law is called after the British ecologist L. R. Taylor (see Taylor[2]). For random networks another power law, that is the power law degree distribution is widely studied (see Barabási and Albert[1]). In this paper a precise mathematical proof is presented that the original Taylor's power law is asymptotically true for the N-stars network evolution model.

We call a graph N-star graph if it has N vertices, one of them is called central vertex, the remaining N-1 vertices are called peripheral vertices, and they are connected to the central one. The N-star network evolution model is the following. At each step either a new N-star is constructed or an old one is activated again. The central weight of a vertex is  $w_1$ , if the vertex was  $w_1$ -times central vertex during the activations. The peripheral weight of a vertex is  $w_2$ , if the vertex was  $w_2$ -times peripheral vertex during activations. In this paper we calculate the mean and the variance of  $w_2$  when  $w_1$  is fixed, and we shall see that the variance function is asymptotically quadratic.

**Keywords:** Taylor's power law, network evolution, N-star

### References

1. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science* 286, no. 5439, 509–512, 1999.
2. Taylor, L.R. Aggregation, variance and the mean. *Nature* 189 (1961), 732-735, doi:10.1038/189732a0.

The research was supported by the construction EFOP-3.6.3-VEKOP-16-2017-00002; the project was supported by the European Union, co-financed by the European Social Fund.

## Design of clinical trials with “time-to-event” end points and under Poisson-gamma enrollment model

Valerii Fedorov

*Innovation Centre, ICON Clinical Research, North Wales, USA*

In clinical trials additionally to observational uncertainties generated by randomness of treatment outcomes, observational errors or by variability between units/subjects we face uncertainties caused by enrollment process that often can be viewed as stochastic processes. The latter makes the amount of information, which can be gained during the trial execution, uncertain at the design stage. The suggested approach guarantees that the information metrics either will be greater than predefined levels with the smallest probability or the average information will be maximized. We illustrate the approach using proportional hazard models with censored observations and enrollment described with the Poisson or more generally the Poisson-gamma process.

**Keywords:** Clinical trials, Poisson-gamma model, Optimal design of experiments

## Healthy Ageing in Czechia

Tomas Fiala<sup>1</sup>, Jitka Langhamrova<sup>2</sup>

*<sup>1,2</sup>Department of Demography, Faculty of Informatics and Statistics, University of Economics, Prague, Praha 3, Czechia,*

For populations of the economically developed countries, long life has become a reality. Currently, life expectancy at birth is around 82 years for women and 76 years for men in the Czechia. At the age of 65, which is the age formally considered as the old age threshold, women have on average more than 20 years and men more than 16 years to live. In addition, even the numbers of the oldest-old increase. Czechia will face an accelerated population ageing process in the next few decades (which will culminate around 2060), continuing decline in mortality, gradual retirement of generations born in the 1970s, and low fertility. The proportion of old-age persons is increasing. About one quarter of people in Czechia in 2040 will be over 65 years old and it will reach 30 % in late fifties. The proportion of persons over 80 years will grow several times until the end of this century. Quality of these years naturally deserves attention, asking whether we “add not only years to life, but above all life

to years". Adding years to life is often referred to as successful ageing. It is important to characterize not only the total lengths of life length of life but also its quality, expressed as the period of life in good health. An important indicator of such type is Healthy Life Years (HLY) indicator. The paper brings actual values of HLY and total life expectancy in Czechia separately for males and females. The limitation prevalence rates will be based of The European Union's Survey on Income and Living Conditions (EU-SILC) survey and European health Interview Survey (EHIS). The comparison and the analysis of the differences between males and females and between healthy and total life expectancies using decomposition method will be presented. The estimate of number of persons with or without activity limitation based on the latest population projection for Czechia until 2050 will be presented.

**Keywords:** Population Ageing, Healths, Healthy Life Years, Czechia

## Construction and Universal Representation of $k$ -Variate Survival Functions

Jerzy Filus<sup>1</sup>, Lidia Filus<sup>2</sup>

<sup>1</sup>Department of Mathematical and Computer Sciences, Oakton Community College, Des Plaines, IL, USA, <sup>2</sup>Department of Mathematics, Northeastern Illinois University, Chicago, IL, USA

In bivariate ( $k = 2$ ) case we consider two arbitrary (in general, not from the same class), fixed, marginal univariate survival functions  $S_1(x_1)$   $S_2(x_2)$ . We propose a method to find a class of corresponding joint survival functions of random vectors  $(X_1, X_2)$ . To find any such a joint survival function, in a general (universal) form, we only need to find a function  $J(x_1, x_2)$  which totally describes stochastic dependences between random variables  $X_1, X_2$ , given in advance the marginals  $S_1(x_1), S_2(x_2)$ . If the corresponding joint probability density exists, then any such a "dependence function"  $J(x_1, x_2)$ , that we call 'joiner', can be found by means of a solution of a derived integral inequality which often reduces itself to the corresponding integral equations which under some simplifying assumptions may become linear. Some 'solutions' for  $J(x_1, x_2)$  are easily obtained so specific new stochastic models can immediately be constructed. All the resulting joint survival functions have the product form  $S_1(x_1) S_2(x_2) J(x_1, x_2)$  which turns out to be universal for all the bivariate survival functions ! The latter fact makes this 'product representation' (as, similarly, in the general  $k$ -variate case)

competitive to the copula methodology. Moreover, the underlying constructions are easier and more efficient than finding copulas especially in application problems. We then turn our attention to (similar) tri-variate and finally to the general  $k$ -variate cases. We show the recurrence transition from  $(k-1)$  to  $k$ -variate case. This type of models may find numerous applications in reliability, econometrics, bio-medical, and other problems.

**Keywords:** probability,  $k$ -variate survival functions and their universal representation, copula alternative, applications

### One step-ahead predictive ability in nested regression models

**S.B. Fotopoulos, S. Lyu, V.K. Jandhyala, A. Kaul**

*Washington State University, Carson College of Business, Washington State University, Department and Management Science, Pullman, WA, United States*

The aim of this study is to reexamine regression-based tests of hypothesis about out-of-sample prediction errors. Forecast accuracy built on a one-step ahead often relates a parsimonious null model to larger models that nests the null model. We reaffirm the asymptotic equivalence between frequently used test statistics for out-of-sample predictive accuracy and  $F$  statistics in the in-sample case. It is shown that under the null hypothesis the larger model introduces just noise into the forecasts values as a result to obtain insignificant results for the extra estimating parameters. If the out-of-sample size is variable, it is shown that the asymptotic test statistics are weakly converge to functional of Brownian motions and for each fixed fraction of sizes their densities are related to Bessel function of third kind. In addition, the asymptotic densities under the alternative are shown to be related to densities of the null hypothesis by a simple convolution operation. Simulation results confirm that the empirical approximated statistics are function of the in-sample ratios, and the terms omitted are shown to be negligible in size. Also, it is revealed that the empirical powers are as efficient as the power of the in-samples case. Finally, goodness-of-fit tests of Kolmogorov-Smirnov and Anderson-Darling type reiterate the merit of asymptotic distributions.

**Keywords:** Model Selection, Overfitting, optimality, out-of-sample variable size

## Poisson regression and change-point analysis

Jim Freeman, Yijin Wen

Alliance Manchester Business School, University of Manchester  
Booth Street East  
Manchester, United Kingdom

Over the years, a diverse repertoire of procedures has evolved for analysing Poisson change-point data. Adding to this collection, the paper presents a new approach based on a Poisson regression formulation. Tested across a range of contrasting datasets, the procedure is shown to perform comparably to mainstream alternatives but with the advantage of being able to distinguish qualitatively between different types of change.

**Keywords:** Change-point, Deviance, Goodness of fit, Poisson regression

## Two-way cross balanced ORDANOVA

Tamar Gadrach

Department of Industrial Engineering and Management, 51 Snunit, ORT Braude College, Karmiel, Israel

Stevens's scales of measurement for *categorical (nominal or ordinal)* variables set the layout of data representation and legitimate operations between these variables. Here, we consider variability of qualitative variable measured according to ordinal scale. The ordinal variation is defined through an appropriate Loss-of-Similarity function applied to Gini Mean Difference measure of variation. Testing null hypothesis, assuming homogeneity between samples drawn from *numerically* described phenomena explained by one (two) factor(s), is usually analyzed by one (two)-way ANOVA. In [1, 2] we introduced the *ORDANOVA* (ORDinal Data ANalysis Of VAriation) procedure as a tool for testing the variation of ordinal data samples explained by only one factor. A generalization presented here focus on searching the variability explanation based on two factors and their possible interaction (*crossed design*) by defining the so-called *two-way ORDANOVA*. Assume factor *A* has *M* levels, factor *B* has *K* levels and *balanced* design. The latter means that per cell, the same amount of *R* independent items were drawn from an infinite population characterized by vector of proportions with *r* categories  $(p_1, p_2, \dots, p_r)$  (i.e.,  $p_i$  denotes the proportion of items belonging to the *i*-th category). We provide a decomposition of the total variation into *intra* component and *inter* component. The inter component is split to the factors *A* and *B* effects



and the  $AB$  interaction. Moreover, we can set appropriate segregation indices (equivalent to  $F$ -statistics) for testing the null hypotheses. In case the null hypothesis is rejected, we also provide unbiased estimators for the variation components. We conclude with some numerical examples.

**Keywords:** Loss-of-Similarity function, Ordinal variation, one-way ORDANOVA, total-variation decomposition, Segregation indices.

**References:**

- [1] T. Gadrich, E. Bashkansky, "ORDANOVA: Analysis of Ordinal Variation", *Journal of Statistical Planning and Inference*, 142, p. 3174-3188, 2012.
- [2] T. Gadrich, E. Bashkansky and R. Zitikis, "Assessing variation: a unifying approach for all scales of measurement", *Quality and Quantity*, 49(3), p. 1145-1167, 2015.

## Robust Minimal Markov Model

**Jesús E. García, V.A. González-López**

*IMECC-UNICAMP, Cidade Universitária, Campinas, Brazil*

In this paper we combine two strategies to improve the final model which represents a set of independent samples. We consider a set of independent samples coming from Markovian processes of finite order and finite alphabet. Under the assumption of the existence of a law that prevails in at least 50% of the samples of the collection, we identify samples governed by the predominant law [1]. The approach is based on a local metric between samples, which tends to zero when we compare samples of identical law and tends to infinity when comparing samples with different laws. The local metric allows to define a criterion which takes arbitrarily large values when the previous assumption about the existence of a predominant law does not hold. By means of this procedure we select the samples which will be used to establish a minimal Markov model from the whole set of samples [2]. We apply this combination of statistical procedures in genomic data.

**Keywords:** Markov process; Robust inference

## Stochastic Profile of Strains of Zika from Tropical and Subtropical Regions

Jesús E. García, V.A. González-López, S.L. Mercado Londoño,  
M.T.A. Cordeiro

*IMECC-UNICAMP, Cidade Universitária, Campinas, Brazil*

We consider a list of 153 strains of Zika (NCBI source, see also [1]) as being a collection of independent samples of stochastic processes related by an equivalence relation (see [2]). The strains are from 12 countries, including Brazil and USA, contributing with most of them. Through an equivalence relationship we build a global profile for all Zika sequences. We compared the global profile with two other profiles built from (i) the 44 strains from Brazil and (ii) the 34 strains from the USA. Given a collection  $\{X_t^j\}_{j=1}^p$  of  $p$  independent discrete time Markov processes with finite alphabet  $A$  and state space  $S=A^o$  ( $o$  is the common memory of the processes), denote by  $P^j(s)$  the probability of the state  $s$  in  $S$  of the process  $j$  and  $P^j(a|s)$  the conditional probability of the process  $j$ , for  $s$  in  $S$  and  $a$  in  $A$ . Consider  $M=\{1,2,\dots,p\}\times S$ , then, the elements  $(i,s)$  and  $(j,r)$  both in  $M$  are equivalent if and only if  $P^i(a|s)=P^j(a|r)$ , for all  $a$  in  $A$ . The equivalence classes define an optimal partition of  $M$ , and it is in relation to this partition that we define the profile of the collection of processes.

**Keywords:** Markov Process; Stochastic equivalence

### References:

- [1] Jesús E. García, V.A. González-López, S.L. Mercado Londoño and M.T.A. Cordeiro. Similarity between Strains of Zika from Tropical and Subtropical Regions (submitted).
- [2] Jesús E. García and S.L. Mercado Londoño. Optimal Model for a Set of Markov Processes. AIP Conference Proceedings (in press)

## Generating a ranking on a set of alternatives from the qualitative assessments given by agents with different expertise

José Luis García-Lapresta<sup>1</sup>, and Raquel González del Pozo<sup>2</sup>

<sup>1</sup> PRESAD Research Group, BORDA Research Unit, IMUVA, Departamento de Economía Aplicada, Universidad de Valladolid, Spain

<sup>2</sup> PRESAD Research Group, IMUVA, Departamento de Economía Aplicada, Universidad de Valladolid, Spain

In this contribution, we consider a group of agents evaluate a set of alternatives through a qualitative scale with the purpose of generate a ranking on the set of alternatives. A decision maker assesses the agents' expertise by means of another qualitative scale. Each of the two qualitative scales is equipped with an ordinal proximity measure that collects the ordinal proximities between the linguistic terms of the scale (see [2] and [1]). To generate the ranking on the set of alternatives, we applied the linguistic voting system introduced and analyzed in [3], with replications of the agents' opinions taking into account their expertise. It is important to emphasize that the ordinal proximity measures used to represent how the linguistic terms of the qualitative scales are distributed are relevant in the process.

**Keywords:** Group Decision-Making, Qualitative Scales, Rankings.

### References

- [1] García-Lapresta, J.L., González del Pozo, R., Pérez-Román, D.: Metrizable ordinal proximity measures and their aggregation. *Information Sciences* 448-449, pp. 149-163, 2018.
- [2] García-Lapresta, J.L., Pérez-Román, D.: Ordinal proximity measures in the context of unbalanced qualitative scales and some applications to consensus and clustering. *Applied Soft Computing* 35, pp. 864-872, 2015.
- [3] García-Lapresta, J.L., Pérez-Román, D.: Aggregating opinions in non-uniform ordered qualitative scales. *Applied Soft Computing* 67, pp. 652-657, 2018.

## Error Detection in sequential laser sensor input

Gwenaël Gatto<sup>1</sup>, Olympia Hadjiliadis<sup>2</sup>

<sup>1</sup>*Department of Mathematics and Statistics, Department of Computer Science, Hunter College, CUNY, New York, USA,* <sup>2</sup>*Department of Mathematics and Statistics, Hunter College, CUNY, New York, USA*

This paper puts forth an online, robust, and low-cost error-detection algorithm to adjust for sensor faults and inaccuracies. The algorithm can detect gradual and sudden sensor slips, and provides measures for real-time corrections. In addition to its reliability, the algorithm does not require any a priori knowledge, nor does it assume the distribution of the data. The runtime is independent of the input size, which is ideal for large volumes of data, and allows for an implementation at the sensor-level.

**Keywords:** CUSUM, Error correction, Hidden Markov models

## Prediction intervals for weighted TAR forecasts

Francesco Giordano<sup>1</sup>, Marcella Niglio<sup>1</sup>

<sup>1</sup>*Department of Economics and Statistics, Università degli Studi di Salerno, Fisciano (SA), Italy*

In this contribution we evaluate the forecast accuracy of a new predictor proposed for the Self Exciting Threshold AutoRegressive (SETAR) model. This model, that belongs to the wide class of the nonlinear time series structures, has been widely studied and applied in the literature for its ability to catch some features often observed in economic, hydrological and financial time series. Instead of the good fitting results, the forecasting performance of the SETAR models has not always given equivalent good results so rising, in some contributions, the need to propose new approaches to generate forecasts. Among them we focus the attention on the predictors obtained as weighted mean of the past observations. In more detail, we consider a weighted mean predictor, that we call weighted SETAR predictor, whose weights are obtained from the minimization of the Mean Square Forecast Errors (MSFE). Even if the “point accuracy” of this weighted predictor has been performed, the study of its distribution and in particular the construction of the prediction intervals (PI) has not yet been faced. Starting from the evaluation that the prediction errors, obtained from the difference between the true future values and the predicted values, follow a nonstandard distribution, in this contribution we focus the attention on different bootstrap methods for dependent data that

allow to construct PI for the weighted SETAR predictor and their coverage is properly compared.

**Keywords:** SETAR model, mean square forecasts error, bootstrap, prediction intervals

## Filling the Gap between Continuous Time Autoregressive Processes and Discrete Observations

Valrie Girardin<sup>1</sup>, Rachid Senoussi<sup>2</sup>

<sup>1</sup>*Laboratoire de Mathématiques Nicolas Oresme, UMR6139,  
Université de Caen Normandie, Campus II, Caen, France*

<sup>2</sup>*INRA, Laboratoire de Biostatistique et Processus Spatiaux (BIOSP), Domaine  
Saint-Paul, Avignon, France*

In theory, continuous time processes may be observed through trajectories on intervals. In practice, data depending on time are mainly collected – regularly or irregularly, sparsely or densely – at discrete observation times, and hence chronologically ordered by integer numbers, finally yielding discrete time sequences. The other way round, many discrete time sequences can be considered as originating from a continuous time process with data sampled at some pertinent time scale. The communication focuses on both discretizing continuous time autoregressive (AR) processes and embedding discrete autoregressive sequences into continuous ones. Both in modeling and for statistical purposes, some compulsory working hypotheses for analyzing data sets lead to the following issue: are these hypotheses usefully preserved at the continuous or at the discrete level? Figuring out how the discretely sampled process inherits the properties of the original continuous version is a classical issue. Conversely, figuring out how properties of a discrete sequence can be used to construct continuous versions yielding back these properties to the ensuing discretized sequence is less classical but of equal interest. The continuous-time AR processes are driven by either Brownian or jump processes, and may have random coefficients depending on time. The innovation of the discrete time processes may be the classical Gaussian, among many other types. In one way, observing the continuous time AR process at discrete times leads the AR dynamics of the discretized process to be characterized. The other way round, AR sequences are embedded, in the almost sure sense, into continuous time

AR processes with the same dynamics. Illustration is provided through many examples and simulation.

**Keywords:** Autoregressive process, Autoregressive sequence, Embedding, Jump Processes, Lévy processes

### **Maximization problem subject to constraint of availability in semi-Markov model of operation**

**Franciszek Grabski**

*Department of Mathematics and Physics, Polish Naval Academy, Gdynia, Poland*

Semi-Markov decision processes theory delivers methods which allow to control an operation processes of the systems. The infinite duration SM decision processes is presented in the paper. The gain maximization problem subject to an availability constraint for the infinite duration Semi-Markov model of the operation in reliability aspect is discussed in the paper. The problem is transformed on some linear programming maximization problem.

**Keywords:** Semi-Markov decision processes, maximization, linear programming

### **Prediction of the 2019 IHF World Men's Handball Championship -- An underdispersed sparse count data regression model**

**Andreas Groll<sup>1</sup>, Jonas Heiner<sup>2</sup>, Gunther Schaubberger<sup>3</sup>, Jörn Uhrmeister<sup>4</sup>**

*<sup>1</sup>Faculty of Statistics, TU Dortmund University, Dortmund, Germany, <sup>2</sup>Faculty of Statistics, TU Dortmund University, Dortmund, Germany, <sup>3</sup>Chair of Epidemiology, Department of Sport and Health Sciences, Technical University of Munich*

*<sup>4</sup>Faculty of Sports Sciences, Ruhr-University Bochum*

In this talk, we compare several different modeling approaches for count data applied to the scores of handball matches with regard to their predictive performances based on all matches from the four previous IHF World Men's Handball Championships 2011--2017: (*underdispersed*) *Poisson regression models*, *Gaussian response models* and *negative binomial models*. All models are based on the teams' covariate

information. Within this comparison, the Gaussian response model turns out to be the best-performing prediction method on the training data and is, therefore, chosen as the final model. Based on its estimates, the IHF World Men's Handball Championship 2019 is simulated repeatedly and winning probabilities are obtained for all teams. The model clearly favors Denmark before France. Additionally, we provide survival probabilities for all teams and at all tournament stages as well as probabilities for all teams to qualify for the main round.

**Keywords:** IHF World Men's Handball Championship 2019, Handball, Lasso, Poisson regression, Sports tournaments

### **Health vulnerability related to climate extremes in Amazonia and the Brazilian Northeast**

**Gilvan Guedes<sup>1</sup>, Pollyane Silva<sup>2</sup>, Maria Helena Spyrides<sup>2</sup>, Claudio Silva<sup>2</sup>, Lara Andrade<sup>2</sup>, Kenya Noronha<sup>3</sup>**

*<sup>1</sup>Demography Department, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, <sup>2</sup>Department of Climate Sciences, Universidade Federal do Rio Grande do Norte, Natal, Brazil, <sup>3</sup>Department of Economics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil*

The Brazilian Amazon and the Brazilian Northeast are the two regions with the highest levels of vulnerability to climate change in the country. While the first is characterized by the largest rainforest in the world and has a very hot and humid climate, the second host one of the largest deserts in the globe. Because of the very low latitudes, these regions are subject to very high temperatures and susceptible to many tropical diseases, such as vector-borne (dengue, malaria, yellow fever), water-borne, and gastrointestinal disease. These diseases are very sensitive to particular climate conditions, such as increase in temperature trend and precipitation concentration. This paper develops a multidimensional index of health vulnerability to climate extremes in Amazonia and the Brazilian Northeast applying the Alkire-Foster method. We use accurate, high quality climate data coupled with health (hospitalizations, disease notification rates, and death rates due to natural hazards), socioeconomic (income, schooling), demographic (young and elderly dependency ratio), and sanitation data for 92 regions. Among the 27 extreme climate indicators produced, we selected 8 for temperature (including temperature range, and maximum of maximum daily temperature) and 7 for precipitation (including number of consecutive dry and wet days). For some climate indicators, we used the

trend coefficient over the 33 years of data and its associated p-value from time series stochastic models in order to select those relevant for the index. Results suggest that 28% of Amazonian regions were deprived in at least 25% of the variables used to create the index, against 8% in the Northeast. The level of health vulnerability varies significantly when homogenous climate zones are taken into account.

**Keywords:** Health Vulnerability, Climate Change, Alkire-Foster Method, Amazonia, Brazilian Northeast

### **Generational differences in health-related quality of life among Brazilian gay men**

**Gilvan Guedes<sup>1</sup>, Samuel Araujo<sup>2</sup>, Paula Ribeiro<sup>3</sup>, Kenya Noronha<sup>4</sup>**

*<sup>1</sup>Demography Department, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, <sup>2</sup>Department of Economics, UFMG, Belo Horizonte, Brazil,*

*<sup>4</sup>Department of Economics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil*

This paper studies the generational differences in health-related quality of life among gay men in Minas Gerais, Brazil. To estimate the health-related quality of life we use the SF-8 (Medical Outcomes Study 8-Item Short-Form Health Survey) and PH-Q9 (Patient Health Questionnaire-9) instruments. Sociodemographic characteristics, LGBT identity measures, and access and utilization of health care services are also considered to differentiate the highly heterogeneous LGBT community. To identify cohort differences, we use a cohort-stratified survey by year of birth and date of markers related to social and institutional LGBT movements in Brazil. Data will be collected through an online platform using the Respondent-Driven Sampling technique, with seeds representing each stratum. Based on the growing literature on health and behavioral differentials within the LGBT community, we hypothesize that older cohorts have lower rates of health-related quality of life than the younger generations due to an increase in acceptance and inclusiveness in recent years. However, as predicted by the minority stress model, we expect that acceptance will be highly asymmetrical across groups due to varying exposure and concrete experiences of violence as a result of prejudice, expectations of rejection, attempts to suppress the LGBT identity, internalized LGBTphobia, and unhealthy coping strategies. Ancillary data collected at pride parades in São Paulo and Belo Horizonte suggest



important differences by cohorts, with older cohorts showing the worse health and quality of life indicators. Their network compositions are also different. Older and younger cohorts show higher diversity in terms of gender identity and sexual orientation, while intermediate cohorts are more homogenous. While the heterogeneity in the older cohorts may be due to their efforts for social acceptability, in the younger it reflects an increase in cultural acceptance.

**Keywords:** Health-related Quality of Life, LGBT, Minority Stress, Mixed-Methods, SF-8, PH-Q9

### **Likelihood-comparison of alternative Markov models incorporating duration of stay**

**Marie-Anne Guerry<sup>1</sup>, Philippe Carette<sup>2</sup>**

*<sup>1</sup>Vrije Universiteit Brussel, Department Business and Technology, Brussels, Belgium, <sup>2</sup>Universiteit Gent, Department of Economics, Ghent, Belgium*

Markov chains are commonly used to model transitions in a system partitioned into categories. In manpower planning models these categories are, for example, job levels or grades in the firm under study. Building a Markov model starts with selecting its states that are assumed to be homogeneous; i.e. the system units in a same state have similar transition probabilities. For systems where the transitions among the categories depend on the duration of stay in the outgoing categories, previous work considered Markov models where the states are subdivisions of the categories into duration of stay intervals, and the more complex semi-Markov models. The present work investigates alternative Markov models for systems where the categories have transition probabilities depending on the duration of stay by selecting the states in different ways: state selection by duration intervals and state selection by duration values. The resulting Markov models are compared based on the likelihood of a set of panel data given the model. For a system with two categories, we prove that the model with states defined by duration values has a better maximum likelihood fit than the base model having the initial categories as states, while this is not the case for the model with states defined by duration intervals under conditions that seem realistic in practice. Although the duration-interval approach is considered in previous studies, the likelihood-comparison is less in favor of this model.

**Keywords:** Markov chain, likelihood, duration of stay, model selection

## Fractional Difference ARFIMA Models for long memory timesies

**Maryam Haghiri**

*Export on statistics, Statistical Centre of Iran, Islamic Republic of Iran, Tehran, Iran*

A class of general models for long memory time series is the fractional differenced models, ARFIMA (p,d,q). This class is a generalization of famous Box – Jenkins ARIMA models, where the parameter d is a real number. The series is stationary and invertible if  $-0.5 < d < +0.5$ . These models are decreasing hyperbolically which is more slower than exponential decay for ARMA. When  $0 < d < 0.5$ , The series have a long memory, and when  $-0.5 < d < 0$  they are unstable or antipersistent.

In this paper, the long memory time series are presented and by defining the predictable memory, we show the methods for choosing the parameters of ARMA adjustment for (0,d,0) along with minimization of prediction variance for one safer ahead forecast.

**Keywords:** Autoregressive processes, Moving average processes, long memory time series, fractional difference methods

## Minimizing Expected Discounted Cost in Queueing Loss Models with Discriminating Arrivals

**Babak Haji**

Sharif University

We consider a queueing loss system with heterogeneous skill based servers and Poisson arrivals. We first assume that each arrival has a vector  $(X_1, \dots, X_n)$  of independent binary random variables with  $X_i = 1$  if server i is eligible to serve that arrival. The service times are exponential with rates depending on the server. Arrivals finding no servers that are both idle and eligible to serve them are lost. Assuming the system incurs a cost of

one unit for each lost customer, our goal is to find the optimal policy for assigning arrivals to idle and eligible servers so as to minimize the expected discounted cost of the system. Later, we generalize our model by considering k server pools where each pool i is eligible to serve arrivals with probability  $p_i$  and all servers within this pool provide service at the same exponential rate.

## Estimation of the relative error in regression analysis under random left-truncation model

**Farida Hamrani**

*Faculty of Mathematics, Department of Probabilities and Statistics,  
U.S.T.H.B., Algiers, Algeria*

In this work, we investigated the relationship between a random covariable and a scalar response which is subject to left truncation by another random variable. Precisely, we use the mean squared relative error as a loss function to construct a kernel estimator of the regression function of this data. We establish the almost sure consistency with rate of the estimator as well as its asymptotic normality. We give also illustrations of the results on simulated data.

**Keywords:** almost sure consistency, asymptotic normality, regression, relative error, left truncation

## Robust Regression in Time Series under Truncated and Censored data

**Benseradj Hassiba<sup>1</sup>, Guessoum Zohra<sup>2</sup>**

*<sup>1</sup>Faculty of Science, Univ. M'hamed Bougara, Algeria, <sup>2</sup>Lab M.S.T.D., Faculty of Math., Univ. Sci. Tech. Houari Boumédiène, Algeria*

We investigate the properties of an M-estimator of the nonparametric regression function based on kernel methods. The strong uniform consistency with rate is established under  $\alpha$ -mixing dependence when the response variable is subject to both random left truncation and right censoring (LTRC). Our results hold with unbounded objective function  $\psi_x$ . A large simulation study of this estimator for one- and bi-dimensional regressor are drawn for fixed and local bandwidth.

**Keywords:** M-estimator, Robust regression, Truncated-Censored data, Strong uniform consistency rate

## **Sequential on-line detection and classification in 3D Computer Vision**

**Olympia Hadjiliadis**

*Department of Mathematics and Statistics, Hunter College, CUNY, NY, USA*

The topic of interest in this talk is the use of on-line statistical sequential detection techniques in automatic 3D image reconstruction.<sup>3</sup> We will begin this presentation by introducing sequential techniques in statistics will stress their importance in applications. In particular, I will contrast the classical hypothesis testing with fixed sample size to sequential decision making and introduce the sequential probability ratio test (SPRT). I will then talk about the problem of quickest detection and introduce the cumulative sum test (CUSUM) and its importance. As an application of the above techniques we will discuss the problem of automatic 3D image reconstruction through laser scan sequential data. We will first apply appropriately tuned CUSUMs to distinguish vertical vs horizontal surfaces. We will then introduce Hidden Markov models to capture vegetation in urban scenes. By applying CUSUMs to detect changes from on Hidden Markov model to another we will be able to identify the beginning of regions of vegetation. By then applying repeated SPRTs, we will be able to identify the ending of these regions. We are thus able to distinguish vertical vs horizontal surfaces as well as regions of vegetation by making use of data sequentially.

**Keywords:** SPRT, CUSUM, Sequential classification of point clouds of urban scenes, 3D Vision

## **Cone distribution functions and quantiles for multivariate random variables**

**Andreas Hamel, Daniel Kostner**

*Free University of Bolzano, Brunico, Italy*

Set-valued quantiles for multivariate distributions with respect to a general convex cone are introduced which are based on a family of (univariate) distribution functions rather than on the joint distribution function. It is shown that these quantiles enjoy basically all the properties of univariate quantile functions. Relationships to families of univariate quantile functions and to depth functions are discussed. Finally, a corresponding Value-at-Risk for multivariate random variables as well as a stochastic

(dominance) order based on quantiles are introduced via the set-valued approach.

**Keywords:** multivariate statistics, set optimization, quantile

## Introducing and evaluating a new multiple component stochastic mortality model

**P. Hatzopoulos, A. Sagianou**

*University of the Aegean, Department of Statistics and Actuarial-Financial Mathematics, Karlovassi, Greece*

This work introduces and evaluates a new multiple component stochastic mortality model. Our proposal is based on a parameter estimation methodology, which aims to reveal significant and distinct age clusters by identifying the optimal number of incorporated period and cohort effects. Our methodology adopts Sparse Principal Component Analysis and Generalized Linear Models (GLMs), which firstly introduced in Hatzopoulos and Haberman (2011), while it incorporates several novelties. Precisely, our approach is driven by the Unexplained Variance Ratio (UVR) metric to maximize the captured variance of the mortality data and to regulate the sparsity of the model with the aim of acquiring distinct and significant stochastic components. In this way, our model gains a highly informative structure in an efficient way, while it is able to designate an identified mortality trend to a unique age cluster. We also provide an extensive experimental testbed to evaluate the efficiency of the proposed model in terms of fitting and forecasting performance over several datasets (Greece, England & Wales, France and Japan), while we compare our results to those of well-known mortality models (Lee-Carter, Renshaw-Haberman, Currie (APC), and Plat). Our model is able to achieve high scores over diverse qualitative and quantitative evaluation metrics and outperforms the rest of the models in the majority of the experiments. Our results advocate the beneficial characteristics of the proposed model and come into agreement with well-established findings of the mortality literature.

**Keywords:** Mortality forecasting, Cohort mortality, Generalized Linear Models, Sparse Principal Component Analysis, Dynamic Linear Regression models, Arima models

## Modeling of the extreme and records values for precipitation and temperature in Lebanon

Ali Hayek<sup>1,2,3\*</sup>, Nabil Tabaja<sup>1,2</sup>, Zaher Khraibani<sup>4</sup>, Samir Abbad Andaloussi<sup>3</sup>, Joumana Toufaily<sup>1,2</sup>, Evelyne Garnie-Zarli<sup>3</sup>, Tayssir Hamieh<sup>1,2\*</sup>

<sup>1</sup>Laboratory of Materials, Catalysis, Environment and Analytical Methods Laboratory (MCEMA), EDST, FS, Lebanese University, Hadath, Lebanon,

<sup>2</sup>Laboratory of Applied Studies for Sustainable Development and Renewable Energy (LEADDER), EDST, Lebanese University, Hadath, Lebanon, <sup>3</sup>

Laboratoire Sol Eau Systèmes Urbains (Leesu) – Université Paris Est-France,

<sup>4</sup>Faculty of Science, Department of Applied Mathematics, Lebanese University, Hadath, Lebanon

Extreme natural phenomena can cause loss of life, damage to infrastructure and very high insurance premiums every year. These phenomena have the potential to reproduce frequently and / or at a very high scale. This is, for example, a heavy rainfall, snowfall and rising temperature. It is therefore important to know about occurrences of such extreme events and their probability of occurrence. The Extreme Value Theory (EVT) is a useful tool to describe the statistical properties of extreme events. Lebanon boasts 225 kilometers (140 miles) of coastline to its west, all of which sits on the eastern Mediterranean Sea, so it is highly exposed natural disasters. Therefore, a database of the temperature and precipitation at Lebanon, chosen as a weather factor, is simulated by two extreme distributions using R program: Weibull for the block maxima (BM) method and General Pareto for the peak over threshold (POT) method to model the tail distributions of temperature and precipitation in Lebanon. Using the theory of record-breaking data, to study the evolution of the temperature and precipitation during 1901-2015. This work predicts the intensity of the next "highest" temperature and computes the probabilities of the waiting time for the future record. In order to study the evaluation of the highest temperature and precipitation to be used in the prediction of return level and the dependency structure based on bivariate extreme value theory and on the conditional probabilities calculated by using the logistic and Husler Reiss models.

**Keywords:** Extreme weather phenomenon, Temperature, Precipitation, Extreme distributions, Block maxima, POT, Return level, Bivariate extreme value, Logistic model, Husler Reiss, Records theory.

## **Births by order and childlessness in the post-socialist countries**

**Filip Hon<sup>1</sup>, Jitka Langhamrova<sup>2</sup>**

<sup>1,2</sup>*Department of Demography, Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic*

This article aims to contribute to women's fertility research in the Czech Republic and other European countries. It focuses on the phenomenon of childlessness and on children by the order of birth.

Female population by number of children ever born is analyzed by age groups. Except the international comparison, this article deals with the projection of the monitored characteristics in the future too.

**Keywords:** Childlessness, birth order, second demographic transition, recuperation.

## **Brand-Level Market Basket Analysis by Conditional Restricted Boltzmann Machines**

**Harald Hruschka**

*Faculty of Economics, University of Regensburg, Regensburg, Germany*

We extend the restricted Boltzmann Machine (RBM) by adding predictors which directly affect probabilities to obtain a conditional restricted Boltzmann Machine (CRBM). The latter differs from the multivariate logit (MVL) model frequently applied in marketing science to analyze shoppers' market baskets by its capability to also reproduce higher order interactions. We compare the CRBM to homogeneous and finite mixture versions of both the RBM and the MVL model. We consider a total of 42 brands across ten food categories. For the MVL and the CRBM models we use the same predictors, namely household attributes (income class, household size) and marketing variables of each brand (shelf price, feature, display, price reduction). Market basket and predictor data originate from a household scanner panel. Models are evaluated by their pseudo log likelihood in a holdout sample of randomly selected households. With respect to this criterion finite mixtures of the MVL are better than the homogeneous MVL. On the other hand, the RBM is superior to the finite mixture MVL model though the former in contrast to the latter does not include predictors. Finite mixture versions of the RBM attain somewhat better pseudo log likelihood values in the estimation

sample, but for the holdout sample become inferior to the less complex homogeneous RBM. That is why we refrain from estimating a finite mixture version of the CRBM. Finally, the homogeneous CRBM turns out to be the overall best model with the highest holdout pseudo log likelihood value.

**Keywords:** Marketing, Market Basket Analysis, Machine Learning, Restricted Boltzmann Machine, Multivariate Logit Model

## **Death, Disease, Failure Prediction: Survival Models vs Statistical Machine Learning/Reduced Order Models**

**Catherine Huber-Carol**

*University Paris Descartes, CNRS (National Scientific Research Center) 8145,  
MAP5, Department of Mathematics and Informatics, 45 rue des Saints-Pères,  
75006 Paris, France*

Several processes allow to deal with the curse of dimensionality imposed by the use of Big Data. We present here several of them that allow to maximize the information left in the data after reducing their dimension. In the particular field of survival analysis, reliability and degradation analysis, several probability models are in use. Flexible parametric models like Weibull and Gamma, semi-parametric models like Cox model and its generalizations, latent variable models like First Hitting Time model, in two versions, parametric and non parametric. But people nowadays tend to present concurrent methods based on machine learning algorithms. We compare here their respective performances for predicting the occurrence of Alzheimer disease. The comparison is done on a French cohort between a regular logistic model and a neural network and deep learning approach.

**Keywords:** Deep Learning, Log-linear models, Neural networks, Projection pursuit, Reduced order models, Singular value decomposition, Survival analysis.



## Solving Rank Aggregation Problems through Memetic Algorithms

Carmela Iorio<sup>1</sup>, Giuseppe Pandolfo<sup>1</sup>, Autilia Vitiello<sup>2</sup>

<sup>1</sup>*Department of Industrial Engineering, University of Naples Federico II, Naples, Italy,* <sup>2</sup>*Department of Physics “Ettore Pancini”, University of Naples Federico II, Italy*

The rank aggregation problem can be encountered in many scientific areas (such as economics, social sciences, computer science, just to cite a few) when the problem is to aggregate a set of individual preferences (rankings or ratings), over a set of alternatives, to find a consensus. The detection of the consensus or median ranking is the identification of the ordering of  $n$  items that best synthesizes the preferences of  $k$  different judges. The median ranking is defined as a ranking that minimizes the sum of distances between itself and all input rankings. The search space of the median ranking is formed by all the possible permutations of the items to be ranked with ties (occurring when a judge assigns the same preference to an object). The distance to be minimized is the Kemeny distance. Since finding a consensus ranking is a Non-deterministic Polynomial-time (NP) hard problem, in the last years Evolutionary Algorithms (EA) are emerging as a suitable methodology to address the complexity of the problem. However, these meta-heuristics are characterized by a slow convergence. To overcome this drawback, in this paper we propose a Memetic Algorithm (MA) to solve the rank aggregation problem. The proposed MA is a combination between genetic algorithms and the stochastic version of the hill climbing search. As shown by a set of experiments performed by exploiting well-known real datasets, our proposal outperforms the evolutionary state-of-the-art algorithms for the rank aggregation problem.

**Keywords:** Rank Aggregation, Consensus Ranking, Kemeny Distance, Evolutionary Algorithms, Memetic Algorithms.

**Acknowledgements:** For Carmela Iorio and Giuseppe Pandolfo this work has been partially supported by the H2020-EU.3.5.4. Project “Moving Towards Adaptive Governance in Complexity: Informing Nexus Security (MAGIC)”, Grant Agreement Number 689669.

## Health status and social activity of men and women at pre- and retirement age in Russia

Alla Ivanova<sup>1</sup>, Elena Zemlyanova<sup>2</sup>, Tamara Sabgaida<sup>3</sup>, Sergey Ryazantsev<sup>4</sup>

<sup>1,4</sup>*Institute of Socio-Political Research, Russian Academy of Sciences, Moscow, Russia; Federal Research Institute for Health Organization and Informatics of Ministry of Health of Russia, Moscow, Russia,* <sup>2,3</sup>*Federal Research Institute for Health Organization and Informatics of Ministry of Health of Russia, Moscow, Russia*

**Significance.** Decision to increase the retirement age in Russia was made against the background of life expectancy growth. However, health status of people of pre-retirement age has been hardly addressed. This issue became the study purpose.

**Materials and methods.** Data on the comprehensive survey of living conditions served as the study information basis (Rosstat, 2016, 4320 respondents aged 45-64). The principal components method was used to describe health status by a set of characteristics (current chronic diseases; disability; limitations in everyday life and the use of rehabilitation means; medical services and their availability; behavioral risk factors). The regression model described relation between health and education, employment, living conditions, income, household composition, leisure and social activities.

**Results.** 8.5% of males and females aged 45 have current chronic conditions, while 4% of males and 3.2% of females have a disability. By the age of 65 the share of chronic patients increases 4 fold, while the share of the disabled – 5.5 fold. Functioning limitations are compensated by rehabilitation means, but their availability depends upon place of residence and social status of people in need. Behavioral risk factors especially smoking remain crucial to the health status. Neither education nor employment reduce risk of chronic diseases but do prevent disability accompanied by cultural and educational leisure and social activity.

**Conclusions.** Feasibility of increasing the retirement age threshold in the context of health status varies considerably across socio-demographic groups.

**Keywords:** health, disability, social activity, retirement age

## **Optimising Group Sequential and Adaptive Clinical Trial Designs: Where Frequentist meets Bayes**

**Christopher Jennison**

*Department of Mathematical Sciences, University of Bath, Bath UK*

The search for efficient group sequential and adaptive designs poses a variety of challenges. It is essential to control the type I error rate, or the familywise error rate when multiple hypotheses may be tested. Efficient trial designs should have good properties over a range of possible scenarios while meeting complex requirements on type I error and possibly on power too. I shall illustrate how frequentist and Bayes methods can be combined to find efficient solutions to clinical trial design problems. The talk will cover early stopping through the use of group sequential tests and sample size modification, as well as related issues in seamless Phase 2-3 trials and adaptive enrichment designs.

**Keywords:** Clinical trials, group sequential, adaptive, frequentist, Bayes

## **A Factor Analysis of Factor Shares, Price Rigidities and the Inflation-Output Trade-Off**

**Christian Jensen**

*Moore School of Business, University of South Carolina, 1014 Greene St, Columbia, USA*

The hypothesis that money affects the aggregate real economy in the short run, despite being neutral in the long run, is one of the most controversial in economics, mainly due to a lack of convincing empirical evidence of short-run non-neutrality, and its relevance at the aggregate level. The present paper studies how inflation can affect real aggregate variables through nominal rigidities that distort price-setting, in order to assess the aggregate relevance of such rigidities empirically. We find that inflation is statistically significant for explaining movements in the income shares of labor, capital and profits, even after controlling for other variables that might generate these co-movements, such as changes in the degree of competition or unionization. These controls are generated through factor analysis, which explains the covariances of the observed variables in terms of the underlying unobservables. Accounting for the observed co-movement between inflation and the income shares without nominal rigidities is difficult, since the income shares are not likely to

impact inflation, or monetary policy, and are independent of most variables and shocks, including those to productivity. Hence, the relationship is evidence of the relevance of nominal rigidities at the aggregate level.

**Keywords:** Price rigidities, Inflation-output trade-off, Phillips curve, Sticky prices, Monetary neutrality, Factor analysis

### **American option pricing under a Markovian regime switching model**

**Lu Jin<sup>1</sup>, Yuji Sakurai<sup>1</sup>, Ying Ni<sup>2</sup>**

*<sup>1</sup>Department of Informatics, University of Electro-communications, Tokyo 182-8585, Japan <sup>2</sup>Mälardalen University, Sweden*

In this research, we consider the pricing of American options when the price dynamics of the underlying risky asset are governed by a Markovian regime switching process. We assume the price dynamics depend on the economy, the state of which transits based on a discrete-time Markov chain. The real state of economy cannot be known directly, but can be partially observed by receiving a signal stochastically related to the real economy. The pricing procedure is formulated using a partially observable Markov decision process, and the optimal strategies which for both put-type and call-type options are investigated. Some properties of the optimal activity regions for each type of option are discussed. Numerical examples are presented to illustrate the results.

**Keywords:** Decision policy, Hidden Markov chain, Optimal strategy, Partially observable Markov Decision Process, Totally Positive of Order 2

### **Revisiting Transitions between Superstatistics**

**Petr Jizba, Martin Prokš**

*Faculty of Nuclear Sciences and Physical Engineering  
Czech Technical University in Prague, Prague, Czech Republic*

This work aims to provide an accurate method for a detection of a transition between Superstatistics. A slight improvement over the currently published method is achieved. Superstatistics framework is briefly recalled and a rather new concept of transition of Superstatistics, introduced by Beck and Xu in 2016, is reexamined. In addition, an original

synthetic model for Superstatistical transition suggested by Beck is discussed. It is shown that its modified version which takes into account a stochastic nature of the transition better reflects empirically observed transitions.

**Keywords:** Superstatistics, Transition of Superstatistics, Monte Carlo simulation, time series

## Generalized T-X family of distributions and their applications

**K. K. Jose, Jeena Joseph**

*St. Thomas College Palai, Mahatma Gandhi University, Kottayam, Kerala,  
Kanichukattu, Pala, Kottayam, Kerala, India*

Generalized families of Statistical Distributions are essential for modeling data sets from a wide range of contexts. In this context we consider the T-X family of distributions and extend them using the Marshall-Olkin transformation. We review the basic theoretical aspects and properties of these distributions. We extend this family to develop a more general family called Marshall-Olkin T-X family. As an illustration we develop the Marshall-Olkin Gumbel Uniform family of distributions and study its properties are explored. The shape properties of the pdf and hazard functions are also examined. The new model is applied on two data sets from industrial contexts and survival analysis. Acceptance sampling plans are developed and minimum values for sample sizes are computed along with operating characteristic functions. Stress strength reliability is obtained and confidence intervals as well as coverage probabilities are computed based on simulation studies. It is also validated with respect to a real data set. Autoregressive minimization processes are also developed for modeling time series data and the sample path properties are explored to illustrate its performance.

**Keywords:** Generalized families, Marshall-Olkin T-X distribution, Time series models, Acceptance sampling plans, Reliability models, stress strength Analysis

## **Distributionally Robust Optimization with Data Driven Optimal Transport Cost and its Applications in Machine Learning**

**Yang Kang<sup>1</sup>, Jose Blanchet<sup>2</sup>, Fan Zhang<sup>2</sup>, Karthyek Murthy<sup>3</sup>**

*<sup>1</sup>Department of Statistics, Columbia University, New York, NY, US,*

*<sup>2</sup>Management Science and Engineering, Stanford University, Stanford, CA, US,*

*<sup>3</sup>Engineering Systems & Design, Singapore University of Technology & Design, Singapore*

Recently, [Blanchet et al. (2016)] showed that several machine learning algorithms, such as square-root Lasso, Support Vector Machines, and regularized logistic regression, among many others, can be represented exactly as distributionally robust optimization (DRO) problems. The distributional uncertainty is defined as a neighborhood centered at the empirical distribution. In this work, we propose a methodology which learns such neighborhood in a natural data-driven way. Also, we apply robust optimization methodology to inform the transportation cost. We show rigorously that our framework encompasses adaptive regularization as a particular case. Moreover, we demonstrate empirically that our proposed methodology is able to improve upon a wide range of popular machine learning estimators.

**Keywords:** Distributionally Robust Optimization, Optimal Transport, Metric Learning, Unbiased Gradient

**References:** J. Blanchet, Y. Kang, and K. Murthy (2016). Robust Wasserstein Profile Inference and Applications to Machine Learning. arXiv preprint arXiv:1610.05627

## **A Joint Modelling Approach in SAS to Assess Association between Adult and Child HIV infections in Kenya**

**Elvis Karanja<sup>1</sup>, Naomi Maina, June Samo**

*<sup>1</sup>University of Nairobi, Nairobi, Kenya*

Recent studies have adopted a joint modelling approach as a more stout technique in studying outcomes of interest simultaneously especially when the interest is in the association between two dependent variables. This has been necessitated by the fact that modelling such outcomes separately often leads to biased inferences due to existing possible

correlations especially in medical studies. This paper demonstrates the application of linear mixed modelling approach using SAS analysis software to evaluate the correlation between adult and child HIV infections for each county in Kenya, while adjusting for several predictors of interest. Using HIV data extracted from the Kenya open data website for the year 2014, we visualize on each county the HIV prevalence on the Kenyan map. High infection incidences are observed for counties located in Nyanza province. We further fit a joint model for the two outcomes of interest using the linear mixed models approach to capture possible correlation between the two outcomes for each county. Results indicate that there is a correlation between infections in adults and children. Further, there is a significant effect of ART coverage, adults and children in need of ART and number of people undergoing testing voluntarily. Researchers or students who have little understanding in application of linear mixed models, both theoretical understanding and practical analysis in SAS as well as application on real datasets, will find this article useful. Findings from this article would interest the health sector, practitioners and other institutions working in HIV related interventions

**Keywords:** Antiretroviral Therapy (ART), HIV, joint modelling, linear mixed model, Repeated measures, SAS

Reference to this paper should be made as follows: Karanja, E., Maina, N., & Samo, J. (2017) 'A Joint Modelling Approach in SAS to Assess Association between Adult and Child HIV infections in Kenya', *Int. J. Data Analysis Techniques and Strategies*,

## **Real time prediction of infectious disease outbreaks based on Google trend data in Africa**

**Elvis Karanja**

*University of Nairobi, Nairobi, Kenya*

New infections with infectious diseases occur quite often in a given susceptible community. However, they do not always lead to an outbreak which would warrant & trigger massive government intervention at the right time. With the advancement in information technology, real-time data collection and dissemination has grown significantly. One of the greatest success stories in real-time disease analytics has been in influenza research using Google flu trends which can predict regional outbreaks of influenza 7-10 days before the center for disease control and prevention

surveillance systems. Other than Google flu trends, Google provides “Google trends” and “Google correlate” which allow a user to input any keywords and obtain data on the number of times Google users searched for such terms. Little has been done with regards to validating Google trends usability in infectious disease prediction in Africa. We propose to bridge this gap by evaluating Google trends data for several infectious diseases, including but not limited to malaria, dengue fever and tuberculosis. To account for internet coverage dynamics, the modeling will be performed for several African countries including Kenya, Uganda, Ethiopia and South Africa.

### **Multivariate Random Sums: Limit Theorems, Related Distributions and Their Properties.**

**Yury Khokhlov<sup>1</sup>, Victor Korolev<sup>2</sup>**

*<sup>1,2</sup>Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, Department of Mathematical Statistics, Moscow Russia*

Random sums are used very often in applied investigations. In one-dimensional case is very popular the model of random sum with index which has geometric distribution. If the summands are independent and identically distributed the limit distribution will be geometric-stable (see the paper by Klebanov 1984). The most popular example are Mittag-Leffler and Linnik distributions. In two paper by Korolev (2016, 2017) the properties of these distributions and their relations with other distributions were investigated in many details.

We consider the multivariate generalization of this problem. This problem is not new, but in papers of other authors it is considered the case of multivariate random sums with common index for all components of the sum. We consider the more general case where the multivariate index of the sum has multivariate geometric distribution with dependent components. We define the notion of multivariate geometric-stable distributions, consider the multivariate analogs of Mittag-Leffler and Linnik distributions, investigate their properties and relations with other distributions using scale mixtures and subordinated processes.

This research is supported by RSCF, project 18-11-00155

**Keywords:** Multivariate random sums, multivariate geometric distribution, scale mixtures, subordinated processes



## Investigating some attributes of periodicity in DNA sequences via semi Markov modelling

Pavlos Koliás<sup>1</sup> and Aleka Papadopoulou<sup>1</sup>

*Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki  
54124, Greece*

DNA segments and sequences have been studied thoroughly during the past decades. One of the main problems in computational biology is the identification of exon intron structures inside genes using mathematical techniques. Previous studies have used different methods, such as Fourier analysis and hidden-Markov models, in order to be able to predict which parts of a gene correspond to a protein encoding area. In this paper, a semi-Markov model is applied to 3-base periodic sequences, which characterize the protein-coding regions of the gene. Analytic forms of the related probabilities and the corresponding indexes are provided, which yield a description of the underlying periodic pattern. Last, the previous theoretical results are illustrated with synthesized and real data from different organisms.

**Keywords:** Semi Markov chains, Periodic patterns, DNA sequences.

## Spatio-temporal Aspects of Community Well-Being In Multidimensional Functional Data Approach

Mirosław Krzyśko<sup>1</sup>, Włodzimierz Okrasa<sup>2</sup>, Waldemar Wołyński<sup>3</sup>

*<sup>1,3</sup>Adam Mickiewicz University, Poland, <sup>2</sup>Cardinal Stefan Wyszyński University in  
Warsaw, Statistics Poland*

This paper has twofold goal due to addressing interconnected methodological and substantive issues involved in modelling both temporal and spatial aspects of the dynamics of local community development and subjective well-being measures. In the first part, the functional data measurement approach - Multivariate Functional Principal Component Analysis (MFPCA) - is applied in a parallel way (independently) to two types of multidimensional measures characterizing, respectively, community and individual (residents') levels of quality (deprivation) and subjective well-being. The MFPCA is an extension of the classic principal component analysis PCA from vector data to functional data (Górecki et al., 2018, 2019) through characterizing units - (local community / commune) or individuals - in terms of many features observed

in many time points and after a smoothing process by a vector of continuous functions. The advantage of the MFPCA over the classic case is to obtain a projection of analysed units into one or two dimensional subspaces using information for the whole period under study, and to divided them into homogenous groups on the basis of the resulting rankings. Having constructed classifications of both local communities (communes) and their residents for the same years (2004 - 2014), the spatial perspective is involved in the second part of the presentation. The space and place-related effects of the community development (deprivation) on the resulting cross-categorization distribution of individuals are evaluated in terms of spatial patterns (autocorrelation and a tendency to clustering) and spatial dependence, spatial regression (Fischer M.M., Getis 2010; Cressie and Wikle, 2011). A multilevel modelling with spatial effect will also be discussed (eg. Arcaya et al., 2012, Okrasa and Rozkrut, 2018). Data come from two sources: (i) measures of local community (communes) development and the relevant covariates are from public statistics, Bank of Local Data (for years 2004, 2008, 2010, 2012, 2014); (ii) subjective well-being measures are based on data from a systematic nation-wide survey Social Diagnosis, carried out in the parallel years. In conclusions, an analytical efficiency of the employed approach is discussed through comparing its outcome with empirical results obtained with the classic PFA-based approach.

#### References:

- Fischer M.M., Getis A. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Springer.
- Cressie N., Wikle Ch. K., 2011. Statistics for Spatio-temporal Data. Wiley.
- Górecki T., Krzyśko M., Waszak Ł., Wołyński W., 2018. Selected statistical methods of data analysis for multivariate functional data, Statistical Papers 59:153–182.
- Gorecki T., Krzyśko M., Wołyński W., 2019. Variable Selection in Multivariate Functional Data Classification. Statistics in Transition ne series (in press).
- Lloyd C.D., 2011. Local Models for Spatial Analysis. CRC Press Taylor & Francis Group.
- Okrasa W., Rozkrut D., 2019. Modelling for Improving Measurement: Strategies for Contextualization of Well-Being. Paper presented at the International Association for Official Statistics /IAOS2018\_OECD Conference Better Statistics for Better Lives. Paris, 19-21 September (2018).

Phillips R., Wong C, 2017. Handbook of Community Well-Being Reserach, Springer

## Mixed Fractional Brownian Motion

Kęstutis Kubilius<sup>1</sup>, Aidas Medžiūnas<sup>1</sup>

<sup>1</sup>*Vilnius University, Institute of Data Science and Digital Technologies  
Department of Methods, Vilnius, Lithuania*

Mixed fractional Brownian motion is a fairly popular research model today. However, there are only a few papers concerned with parameter estimation in the mixed model. For example, although these processes are predominated by the Wiener process, but the presence of fBm calls for the necessity of estimating the Hurst parameter. We consider mixed stochastic differential equation:

$$X_t = x_0 + \int_0^t f(s, X_s) ds + \int_0^t g_1(s) dW_s + \int_0^t g_2(s) dB_s^H,$$

where  $H \in [0,1]$ .

We investigate the asymptotic behaviour of the first and second quadratic variations of the solution, suggest several Hurst index estimates based on these variations and prove their strong consistency.

**Keywords:** Mixed fractional Brownian motion, quadratic variations, asymptotic analysis

## Optimal collection of two materials from N ordered customers with stochastic continuous demands

Epaminondas G. Kyriakidis<sup>1</sup>, Theodosis D. Dimitrakos<sup>2</sup>,  
Constantinos C. Karamatsoukis<sup>3</sup>

<sup>1</sup>*Department of Statistics, Athens University of Economics and Business, Athens, Greece,* <sup>2</sup>*Department of Mathematics, University of the Aegean, 83200, Samos, Greece,* <sup>3</sup>*Department of Military Sciences, Hellenic Military Academy, Vari, Attica, Greece*

A vehicle starts its route from a depot and visits N ordered customers in order to collect from them two materials (Material 1 and Material 2). Each customer has either Material 1 or Material 2. The quantity of the material that each customer possesses is a continuous random variable with known distribution. The type of Material and its actual quantity are

revealed when the vehicle arrives at a customer's site. The vehicle has two compartments (Compartment 1 and Compartment 2) with same capacity. Compartment 1 is suitable for loading Material 1 and Compartment 2 is suitable for loading Material 2. If a compartment is full, it is permissible to load the corresponding material into the other compartment. In this case a penalty cost is incurred that is due to some extra labor for separating the two materials when the vehicle returns to the depot to unload the materials. The travel costs between consecutive customers and between a customer and the depot are known and satisfy the triangle property. The vehicle may interrupt its route and return to the depot to unload the materials. As soon as the material of the last customer has been collected, the vehicle returns to the depot to terminate its route. Our objective is to find a routing strategy that minimizes the total expected cost for serving all customers. The assumption that the customers are ordered enables us to solve the problem by developing a suitable stochastic dynamic programming algorithm. It is shown that the optimal routing strategy has a specific structure. Numerical results are obtained by discretizing the state space.

**Keywords:** vehicle routing problem, stochastic dynamic programming, continuous demand

### Identifying the characteristics influencing the mathematical literacy in Spanish students

**Ana María Lara-Porras, María del Mar Rueda-García, David Molina-Muñoz**

*Department of Statistics and Operational Research, University of Granada,  
Spain*

The average score in mathematics obtained in the PISA 2015 tests by the Spanish students is below the OECD average and that of the EU. In addition, results show important differences in the students' performance between the regions in which Spain is divided into. The aim of this work is to investigate the factors that contribute to the mathematical performance of the Spanish students in the PISA 2015 tests. We analysed variables related to the students and their family-background characteristics, to the schools and to the regions where the schools are located. Due to its hierarchical organisation, where each student belongs to a school and, in turn, each school is located in a region, we considered a multilevel regression model with three levels (students, schools and regions) to

analyse the data. Our results indicated that most of the variables with a significant influence on the students' mathematical performance were characteristics of the students themselves. These variables were the female gender, the grade repetition and the immigrant status (in a negative sense) and the pre-primary schooling and the economic and sociocultural status (in a positive sense). Some variables of the school level such as the index of school responsibility for curriculum and assessment and the total school enrolment emerged as significant, both in a positive sense. Finally, the regional unemployment rate and the GDP of the regions also influenced the students' performance.

**Keywords:** Educational assessment, Mathematical performance, Multilevel analysis, Performance factors, PISA

## I-Delaporte process and applications

Meglana D. Lazarova<sup>1</sup>, Leda D. Minkova<sup>2</sup>

<sup>1</sup>*Faculty of Applied Mathematics and Informatics, Technical University of Sofia, Sofia, Bulgaria,* <sup>2</sup>*Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria*

In this paper we introduce a mixed Polya-Aeply process with shifted gamma mixing distribution and call it an Inflated-parameter Delaporte process (I-Delaporte process). We derive the probability mass function, moments and some basic properties. Then we define a process as a pure birth process and derive differential equations for the probabilities. As application, we consider a risk model in which the claim counting process is the defined I-Delaporte process. For the defined risk model we derive the joint distribution of the time to ruin and the deficit at ruin as well as the ruin probability. We discuss in detail the particular case of exponentially distributed claims.

**Keywords:** Mixed distributions, Pure birth process, Delaporte process, Ruin probability

**Acknowledgement.** The authors are partially supported by project "Stochastic and simulation methods in the field of medicine, social sciences and dynamic systems" funded by the National Science Fund of Ministry of Education and Science of Bulgaria (Contract No DN 12/11/20 Dec. 2017).

## Mobile learning for training bioinformatics in the connected world

Taerim Lee<sup>1</sup>, Juhan Kim<sup>2</sup>

<sup>1</sup>Department of Bioinformatics and Statistics, Korea National Open University/86 Daehak ro Jongro gu, Seoul Korea, <sup>2</sup>Medical College, Seoul National University/103 Daehak ro, Seoul, Korea

This project promotes the implementation of mobile learning initiative in Bioinformatics Training & Education Center (BITEC) supported from Korean Ministry of Health and Welfare. It is 5 years projects co-work together Seoul National University Medical College. We build up KNOU OER LMS system for training nationwide medical doctors and data scientist too. Using ICT the world becoming closely connected and mobile will be an easy accessible educational media for training bioinformatics and data analysis for medical doctors in the era of big data. It was estimated that 95% of the global population living in an area covered by at least a basic mobile cellular network. Global learner have access to the internet and it is expected to continue to rise as more and more open and distance learners, LLL learners come online. The rapid growth in broadband access and usage, driven by mobile broadband technologies, has fostered the development of a mobile learning for training open & distance connected learner. The high penetration rates of mobile phone subscriptions and the rapid growing of handheld users transform higher education through digitally supported learning & teaching for learner. The BITEC m-Learning initiative focuses on introducing Bioinformatics, Medical Informatics Health Informatics and Data Analysis using handheld devices to be made easily accessible for medical doctors on the field and open up ubiquitous learning environment. Lesson learned from this initiative is that the mobile e-Book could be the most affordable, accessible and flexible educational media. Consequently, more accessible tertiary education will meet the demands of population that did not have the time and place for such learning.

**Keywords:** mobile learning, BITEC, ODL, Bioinformatics

## Data Fraud and Inlier Detection

**H.-J. Lenz, W. Kössler**

*Freie Universität Berlin, Germany*

Data fraud is multi-facette domain and includes data manipulation and fabrication besides of data theft and plagiarism. We are concerned with the first two areas, and, especially, with the detection of inliers. In our context, simply speaking inliers are numerical values from a second distribution inserted into the target distribution due to fraud. Tricksters often avoid outliers due to the risk of being detected and prefer values near to the mean of the target distribution. This evidence gives raise to applying a likelihood-ratio test for separating two mixed distributions. The methodology will be explained and the LR-test performance illustrated.

## Balancing Covariates in Regression Discontinuity Designs

**Shuangning Li**

**(Joint work with Stefan Wager, Stanford University)**

*Department of Statistics, Stanford University*

In applied work, it is common to control for auxiliary covariates when deploying a regression discontinuity design (RDD). Although such covariate adjustments are not strictly required for identification, they can be used to improve both precision and robustness. In this paper, we introduce a new way of using auxiliary covariates in RDDs motivated by balancing estimators for causal inference. Our approach can be seamlessly integrated into a rich variety RDDs, including multivariate problems with irregularly shaped treatment boundaries (e.g., geographic RDDs). In our formal analysis, following recent work by Calonico, Cattaneo, Farrell, and Titiunik (2019), we do not make any parametric assumptions on the way in which the conditional response surface depends on covariates. We then show that the amount by which our method improves precision depends on the extent to which the contribution of the covariates can be approximated by a linear function.

## **Using the Developing Countries Mortality Database (DCMD) to Probabilistically Evaluate the Completeness of Death Registration at Old Ages**

**Nan Li<sup>1</sup>, Hong Mi<sup>2</sup>**

*<sup>1</sup>Population Division, Department of Economic and Social Affairs, United Nations, 2UN Plaza, Room DC2-1938, New York, USA, <sup>2</sup>Corresponding author, School of Public Affairs, Zhejiang University, Room 265#, Mengminwei Building, Zijing'gang Campus, Hangzhou, Zhejiang, P. R. China*

*The work on this paper was supported by the Nature Science Foundation of China (NSFC) project (NO.71490732), NSFC project (NO.71490733), Zhejiang SocialScience Planning Project Key Program (NO.17NDJC029Z), and Nature Science Foundation of Zhejiang project (NO.LZ13G030001)*

As the ratio of registered deaths to total deaths, the deterministic completeness of death registration (DR) cannot be exactly 1 in practice. Consequently, it is impossible to use deterministic completeness to check whether a DR is complete, which is a problem for developed countries. We propose a probabilistic completeness whose samples are the values of deterministic completeness. When the difference between 1 and the mean of probabilistic completeness is statistically insignificant, the DR is probabilistically complete. But using intercensal population change to estimate deaths and deterministic completeness is still an issue, because it requires unrealistic assumptions about migration and census error. Focusing on old age and the level of mortality rather than the number of death, the effects of migration and census error are largely reduced in the Developing Countries Mortality Database (DCMD, [www.lifetables.org](http://www.lifetables.org)), which is used to provide applications of the probabilistic evaluation in this paper.



## Gaussian Limits for Multichannel Networks with Input Flows of General Structure

Hanna V. Livinska<sup>1</sup>, Eugene O. Lebedev<sup>2</sup>

<sup>1</sup>Taras Shevchenko National University of Kyiv, Applied Statistics Department, 64/13, Kyiv, Ukraine, <sup>2</sup>Taras Shevchenko National University of Kyiv, Applied Statistics Department, 64/13, Kyiv, Ukraine

A multi-channel queueing network with an input flow of general structure arrived into each node is considered. Each node operates as a multi-channel queueing system. Once the service is completed at a node, the customer is transferred to another node or it leaves the network with correspondent probabilities. Input flows into the different nodes can be interdependent. Service times of customers are independent random values with exponential type distributions. The multi-dimensional service process is introduced as the number of customers at network nodes. We consider processing such a network under certain heavy traffic conditions. Heavy traffic assumptions on network parameters are formulated. It is proved that in this case the multi-channel service process converges to a Gaussian process in the uniform topology. Correlation characteristics of the Gaussian process are written via network parameters in an explicit form. A network with nonhomogeneous Poisson input flow is studied as a particular case of the general model, correspondent Gaussian limit process is built.

**Keywords:** Multichannel Queueing Network, Heavy Traffic, General Input Flow, Gaussian Approximation

## A cluster analysis of multiblock datasets

Fabien Llobell<sup>1,2</sup>, Véronique Cariou<sup>1</sup>, Evelyne Vigneau<sup>1</sup>, Amaury Labenne<sup>2</sup>, El Mostafa Qannari<sup>1</sup>

<sup>1</sup>StatSC, ONIRIS, INRA., Nantes, France, <sup>2</sup>Addinsoft, XLSTAT. Paris, France

CLUSTATIS is a general procedure of cluster analysis of a collection of datasets. It is based on the optimization of a criterion and consists of a hierarchical cluster analysis and an iterative algorithm akin to K-means. Its interest is discussed and illustrated in sensory and preference studies.

### Introduction

We propose a cluster analysis approach of multiblock datasets. This approach consists in an extension of the CLV method [1]. It aims at

minimizing a criterion which reflects the fact that we are seeking homogeneous clusters of datasets. More precisely, the datasets in each cluster are assumed to be highly related to a latent configuration which is determined by means of the STATIS method [2]. More precisely, if we denote by  $X_1, \dots, X_m$  the datasets at hand, which are assumed to be centered. We compute the scalar products matrices:  $W_1 = X_1 X_1^T, \dots, W_m = X_m X_m^T$ , and we seek to minimize the following criterion:

$$\sum_{k=1}^K \sum_{i \in G_k} \|W_i - \alpha_i W^{(k)}\|^2$$

where  $\alpha_i$  ( $i=1, \dots, m$ ) are scalars to be determined,  $K$  is the number of groups,  $G_k$  is the  $k^{\text{th}}$  group of datasets and for ( $k=1, \dots, K$ ),  $W^{(k)}$  is the compromise of the group  $G_k$ . It turns out that the minimization of this criterion leads to determining  $W^{(k)}$  as the STATIS compromise of the datasets in group  $G_k$ . The general procedure of cluster analysis is called CLUSTATIS and consists of two complementary strategies. The first strategy consists in an iterative algorithm akin to the K-means algorithm. The second strategy consists in a hierarchical cluster analysis. Both strategies aim at optimizing the same criterion and, in practice, complement each other. More precisely, the hierarchical cluster analysis can help selecting the appropriate number of clusters and provides a starting partition of the datasets that can be improved by means of the iterative algorithm. We also discuss extensions of the method of analysis. As an illustration, we consider two case studies pertaining to sensory and consumer studies.

**Keywords:** Cluster analysis, STATIS, Multiblock datasets, Sensory analysis

**References:**

- [1] Vigneau and Qannari. Clustering of variables around latent component (2003). *Communications in Statistics – Simulation and Computation*, 32(4):1131–1150.
- [2] Lavit, C., Escoufier, Y., Sabatier, R., & Traissac, P. (1994). The act (statist method). *Computational Statistics & Data Analysis*, 18(1), 97–119.

## Properties of the Hardlims\*Tansig Model of the Statistical Neural Network

**Olamide O. Ilori, Christopher G. Udomboso**

*University of Ibadan, Department of Statistics, Ibadan, Oyo 20005, Nigeria*

This study analytically derived a heterogeneous transfer function using the symmetric hard limit as well as the hyperbolic tangent sigmoid transfer functions from homogeneous statistical neural networks. The methodology used is the statistical neural network model proposed by Anders in 1996. A convoluted form of the artificial neural network function given by Udomboso in 2013 using product convolution was employed. Moreover, the distributional properties of the resulting heterogeneous statistical neural network were investigated, and the mean as well as the variance were shown to exist. Data were generated from the normal distribution with mean of 5 and variance of 1, and were used to demonstrate the parent and derived models. A fixed hidden neuron was used in the models. Analyses were computed using MATLAB R2015a at 1000 iterations. Mean and variance were computed for each prediction and their generated errors. Also computed is their network information criterion. Results showed that for the predicted values, the heterogeneous statistical neural network model with symmetric hard limit and hyperbolic tangent sigmoid transfer function had the least mean and variance, while in the case of the generated error, the homogeneous statistical neural network model with hyperbolic tangent sigmoid transfer function had the least mean and variance. Model selection based on the network information criterion showed that the heterogeneous statistical neural network model with symmetric hard limit and hyperbolic tangent sigmoid transfer function is the better preferred model, while the hyperbolic tangent sigmoid is the least preferred.

**Keywords:** Transfer function, convolution, probability distribution, mean, variance

### **Properties of the extreme points of the probability density distribution of the Wishart matrix**

**Karl Lundengård<sup>1</sup>, Asaph Keikara Muhumuza<sup>1,2</sup>, Sergei Silvestrov<sup>1</sup>, John Mango<sup>3</sup>, Godwin Kakuba<sup>3</sup>**

*<sup>1</sup>Division of Applied Mathematics, The School of Education, Culture and Communication, Mälardalen University, Västerås, Sweden, <sup>2</sup>Department of Mathematics, Busitema University, Box 238 Tororo, Kampala, Uganda, East Africa, <sup>3</sup>Department of Mathematics, College of Natural Sciences, Makerere University, Kampala, Uganda, East Africa*

We will examine some properties of the extreme points of the probability density distribution of the Wishart matrix using properties of the Vandermonde determinant and show examples of applications of these properties.

**Keywords:** Wishart matrix, Vandermonde determinant, extreme points

### **Comparison of parametric models applied to mortality rate forecasting**

**Karl Lundengård<sup>1</sup>, Samya Suleiman<sup>2</sup>, Hisham Sulemana<sup>1</sup>, Milica Rancic<sup>1</sup>, Sergei Silvestrov<sup>1</sup>**

*<sup>1</sup>Division of Applied Mathematics, UKK, Mälardalen University, Västerås, Sweden, <sup>2</sup>Department of Mathematics, College of Natural and Applied Science(CONAS), University of Dar es Salaam, Dar es Salaam, Tanzania*

Mortality rates of a group of humans is very important to consider when determining the overall well-being of the group or planning this like pensions or life insurance. In some situations, it is desirable to have a simple mathematical model for the mortality rate. Many such models have been suggested but there are very little systematic comparisons of the different models available in literature. In this paper we will examine and compare the properties of a selection of models from literature. The models will be fitted to measured mortality rates from different countries and the resulting mortality rates will then be used to predict future mortality rates and the advantages and disadvantages of the different models will be discussed.

**Keywords:** mortality rate, non-linear curve-fitting, forecasting

## Particle filter impoverishment under an urn model perspective

**Rodi Lykou, George Tsaklidis**

*Department of Statistics and Operational Research, School of Mathematics,  
Aristotle University of Thessaloniki, Greece*

The quest of sample impoverishment in particle filtering, namely the phenomenon when all particles of the filter end up to take the same value, is crucial for the precision of the estimates for the hidden states at a certain time point. Thus, the probability that every particle be able to take one out of  $m$  different values in  $n$  time steps after the beginning of the process seems interesting. This view of the problem is relative to an urn model with balls of  $m$  different colours. In this study, special cases of the aforementioned model are examined.

**Keywords:** particle filter; sample impoverishment; urn model

## Bayesian model for mortality projection: evidence from Central and Eastern Europe

**Justyna Majewska, Grazyna Trzpiot**

*University of Economics in Katowice, Katowice, Poland*

Multi-population models for forecasting mortality rates have been the major focus of many authors since the seminal work by Lee and Li (2005). Models are typically based on the assumption that the forecasted mortality experiences of two or more related populations converge in the long run. In much of the existing stochastic-mortality works a two-stage approach is taken into account (model fitting and parameter estimation). Since some shortcomings of the two-level hierarchical procedure are known (e.g. incoherence), the single-step Bayesian approach is adopted. The main reasons of using this approach are (Cairns et al, 2011): 1) it helps to take account of parameter uncertainty in a natural and coherent way, 2) the careful specification of a limited number of prior distributions helps to avoid unreasonable model parametrisations and 3) it allows to deal simply and effectively with small populations, possibly with ! substantial quantities of missing data. In this work we implement the Bayesian approach to model and project mortality rates for more than two populations. Populations

from Central and Eastern Europe with similar socio-economic environments are selected. The Lee-Carter model is used to explore and discuss the technicalities of Bayesian mortality model.

**Keywords:** mortality models, Bayesian approach, multi-population

## Itô Type Bipartite Fuzzy Stochastic Differential Equations with Osgood condition

**Marek T. Malinowski**

*Institute of Mathematics, Cracow University of Technology, Kraków, Poland*

We consider bipartite fuzzy stochastic differential equations [9,10] as a tool in modeling dynamical systems operating in random and fuzzy environment. Such equations possess integrals on both sides and both sides are significant. These equations cannot easily be reduced to the equations with only one side. The difficulty lies in the issue of the difference of fuzzy sets. Such difference may not exist. In addition, each side of the equation has a different effect on the properties of solutions. This is about behaving a function whose values are the diameter of the solution at time  $t$ . The right-hand side drives the increase in diameter while the integrals on the left force the diameter to decrease. The bipartite equation combines two different types of equations previously investigated in the literature, i.e., equations with increasing fuzziness [1-6] and equations with decreasing fuzziness [7,8]. In the communication we will consider a problem of existence of solution under condition which is weaker than Lipschitz condition used in [9,10].

**Keywords:** Fuzzy stochastic differential equation, modelling in fuzzy and random environment

### References:

1. Malinowski, M. T., "Strong solutions to stochastic fuzzy differential equations of Itô type", *Math. Comput. Modelling* 55, 918-928 (2012).
2. Malinowski, M. T., "Some properties of strong solutions to stochastic fuzzy differential equations", *Inform. Sciences* 252, 62-80 (2013).
3. Malinowski, M. T., "Fuzzy and set-valued stochastic differential equations with local Lipschitz condition", *IEEE Trans. Fuzzy Syst.* 23, 1891-1898, (2015).
4. Malinowski, M. T., "Set-valued and fuzzy stochastic differential equations in M-type 2 Banach spaces", *Tohoku Math. J.* 67, 349-381, (2015).

5. Malinowski, M. T., "Fuzzy stochastic differential equations of decreasing fuzziness: approximate solutions", J. Intell. Fuzzy Syst. 29, 1087-1107, (2015).
6. Malinowski, M. T., Agarwal, R. P., "On solutions to set-valued and fuzzy stochastic differential equations", J. Franklin Inst. 352, 3014-3043, (2015).
7. Malinowski, M. T., "Stochastic fuzzy differential equations of a nonincreasing type", Commun. Nonlinear Sci. Numer. Simulat. 33, 99-117 (2015).
8. Malinowski, M. T., "Fuzzy stochastic differential equations of decreasing fuzziness: non-Lipschitz coefficients", J. Intell. Fuzzy Syst. 31, 13-25, (2016).
9. Malinowski, M. T., "Bipartite fuzzy stochastic differential equations with global Lipschitz condition", Math. Probl. Eng. vol. 2016, 13 pages, (2016).
10. Malinowski, M. T. „On Bipartite Fuzzy Stochastic Differential Equations”, IJCCI Vol. 2: FCTA, eds. Juan Julian Merelo (et al.), SCITEPRESS - Science and Technology Publications, Lda., 109-114, (2016).

## **Asymptotics of Implied Volatility in the Gatheral Double Stochastic Volatility Model**

**Mohammed Albuhayri<sup>1</sup>, Anatoliy Malyarenko<sup>1</sup>, Sergei Silvestrov<sup>1</sup>,  
Ying Ni<sup>1</sup>, Christopher Engström<sup>1</sup>, Finnan Tewolde<sup>1</sup>, Jiahui Zhang<sup>1</sup>**

<sup>1</sup>*Division of Applied Mathematics, Mälardalen University, Västerås, Sweden*

The double-mean-reverting model by Gatheral (2008) is motivated by empirical dynamics of the variance of the stock price. No closed-form solution for European option exists in the above model. We study the behavior of the implied volatility with respect to the logarithmic strike price and maturity near expiry and at-the-money. Using the method by Pagliarani and Pascucci (2017), we calculate explicitly the first few terms of the asymptotic expansion of the implied volatility within a parabolic region.

**Keywords:** Double-mean-reverting, European option, implied volatility, asymptotic expansion

### **References:**

- [1] Gatheral, J. *Consistent Modelling of SPX and VIX Options*. Bachelier Congress, London, 2008.

- [2] Pagliarani, S., Pascucci, A. The exact Taylor formula of the implied volatility. *Finance Stoch.* **21** (2017), no. 3, 661–718.

## Latent class detection in Latent Growth Curve Models

**Katerina M. Marcoulides, Laura Trinchera**

*NEOMA Business School, Rouen, France*

Latent growth curve modeling is frequently used in social and behavioral science research to analyze complex developmental patterns of change over time. Although it is commonly assumed that individuals in an examined sample will exhibit similar growth trajectory patterns, there can be situations where typological differences in development and change are present. In such instances, it is important to assume that the underlying population consists of a fixed but unknown number of groups or classes, each with distinct growth trajectories. Because group membership is not known and no observed variable is available to identify homogenous groups, group membership must in some manner be inferred from the data. We propose a new approach to growth mixture modeling where the number of growth trajectories is determined directly from the data by algorithmically grouping or clustering individuals who follow the same estimated growth trajectory based on an evaluation of individual case residuals. The identified groups are assumed to represent latent longitudinal segments or strata in which variability is characterized by differences across individuals in the level (intercept) and shape (slope) of their trajectories and their corresponding individual case residuals. The illustrated approach algorithmically enables the data to determine both the number of groups and corresponding trajectories. The approach is illustrated using both empirical longitudinal and simulated data.

**Keywords:** Unobserved Heterogeneity, SEM, REBUS

## Properties of patterns in a semi-Markov chain

**Brenda Ivette Garcia Maya<sup>1</sup> and Nikolaos Limnios<sup>2</sup>**

<sup>1</sup> Sorbonne University, Université de technologie de Compiègne,

LMAC Laboratory of Applied Mathematics of Compiègne

CS 60319-57 avenue de Landshut COMPIEGNE CEDEX, 60203 FRANCE

A pattern in a sequence has different properties, for instance, its probability of apparition, its mean waiting position or the rate of its occurrence. Nevertheless, identifying a pattern in a chain formed by large



sequence of symbols can not be done by simple inspection. To carry out this assignment in short time spans several mathematics models have been proposed under the assumption that the chain is described by a discrete Markov process. Even if Markov discrete process describes properly and straightforward a sequence of symbols, the main drawback in the Markov hypothesis is that they can not take into account general distributions in the sojourn time in a state, the sojourn time in a state must be governed by the geometric distribution (in a discrete chain), in contrast discrete-time semi-Markov process generalize the Markov hypothesis allows the distribution function in a state be any one. For this reason, we present a mathematical model which gives the principal properties of a pattern under the semi-Markov hypothesis. To this end we use the auxiliary prefix and backward chain. We compute the probability that a pattern occurs for the first time after  $n$  symbols. The model and algorithm proposed can be applied in many areas, for example communications, informatics, biology, linguistics, etcetera. To exemplify this, the model is tested in a particular pattern on a bacteriophage DNA sequence.

**Keywords:** semi-Markov chain, first occurrence of a word, prefix process, backward position process, DNA analysis.

## **Generalized First Passage Time Method for the Estimation of the Parameters of the Stochastic Differential Equation of the Black-Scholes Model**

**Samia Meddahi, Khaled Khaldi**

*Boumerdes University, Avenue de l'indépendance, Boumerdes, Algeria*

The parameters estimation is one of dynamic models problems in many scientific fields, particularly in finance. This paper presents:

- 1 - the First Passage Time (FPT) method generalized for all Passage Times (GPT), based on the inverse Gaussian law and the first passage time, in order to estimate the parameters of the Black-Scholes model,
- 2 - the results of estimated parameters in a simulated time series, then the computed errors and forecast.

**Keywords:** First passage time, density of transition, Black-Scholes model

## Skorokhod embeddings in Brownian Motion and applications to Finance

**Isaac Meilijson**

*Tel-Aviv University*

Skorokhod (1965) showed that for every distribution with mean zero and finite variance there exist integrable stopping times  $\tau$  in SBM  $B(\cdot)$  such that  $B(\tau)$  has the given distribution. The consecutive application of this idea embeds random walks, and martingales more generally (Doebelin (1940), Dubins & Schwarz (1965), Monroe (1972)), into SBM. The history of Skorokhod embeddings will be briefly outlined, with emphasis on Chacon & Walsh (1976), Dubins (1968) and Azéma & Yor (1978). After describing their role in the study of risk aversion, a weaker notion selective risk aversion (Landsberger & M. 1990) will be shown to be characterized by Skorokhod embeddability in the Azéma martingale, to be described. For a random walk  $S_n$  with positive drift there is (the adjustment coefficient or Aumann-Serrano (2006) index)  $\alpha > 0$  such that  $\exp \{-\alpha S_n\}$  is a martingale. M. (2008) Skorokhod-embedded this martingale in the corresponding exponential transform of a suitable Brownian Motion to infer on random walk Lundberg-type approximations and inequalities related to global minimum and drawdown behavior, from the corresponding answers for Brownian Motion. Time permitting, various other applications to Finance of the adjustment coefficient via Skorokhod embeddings will be described.

## Kernel estimator regression in censored and associated models

**Nassira Menni<sup>1</sup>, Abdelkader Tatachak<sup>2</sup>**

<sup>1</sup>*Faculty of Science, University Algiers I, Department of Mathematics and Computer Science (MI), Algiers, Algeria,* <sup>2</sup>*Lab. MSTD, Faculty of Mathematics, BP 32, El Alia, 16111, University of Science and Technology Houari Boumediene, Algiers, Algeria*

In this work, we are interested to the nonparametric estimator of the regression function in the case of the right censoring data and presenting a form of dependence called association. We aim at establishing some asymptotic properties of the kernel estimator introduced by Guessoum and Ould saïd (2008) while taking a dependency framework for the data. We give the rate of almost sure uniform convergence of the kernel

regression estimator when the data are right censoring and associated. Moreover, we show that the suitably standardized estimator is asymptotically normal. We give also illustrations of our results on simulated data.

**Keywords:** Association, right censoring, almost sure uniform convergence, Kaplan-Meier estimator, kernel estimator, nonparametric regression, asymptotic normality

### Discrete Time Risk Model

**Leda D. Minkova**

*Sofia University, Sofia, Bulgaria*

In this paper we consider a discrete time risk model. We suppose that the counting process is a compound binomial process with geometric compounding distribution. The resulting process, called I-Binomial process is a discrete analog of the Polya-Aeppli process. It is a discrete time stationary renewal process with geometrically distributed inter-arrival times. For the corresponding risk model, we analyze the ruin probability and consider the case of geometrically distributed claims.

**Keywords:** Discrete time risk model, I-Binomial process, ruin probability

**Acknowledgement.** The research was partially supported by Grant DN12/11/20.dec.2017 of the Ministry of Education and Science of Bulgaria.

### Multichannel sequence analysis to identify patient pathway

**Elisabeth Morand**

*Institut National d'Etudes Démographiques (Ined), Paris, France*

The data of the national French Health Insurance system : the SNIIRAM database (Système national d'information inter-régimes de l'Assurance maladie) are a very rich source of information on the drugs delivered, their type, the timing and the quantity. The time where drugs are taken and the diseases treated are not available. Patient pathway is defined as drugs sequence, i.e, an ordered list of successive of drugs delivered. For a given disease the list of drugs is finite. In our case, the treatment consists in a set of drugs delivered for first treatment intention or for furthermore. The aim is to identify patterns of first treatment intention. A multichannel qualitative Harmonic Analysis is performed to cluster patient pathways

using both drugs succession and administrative information. Rules obtained from the clustering is used to determinate a sample of patterns of first treatment intention on a longitudinal cohort of 500 000 beneficiaries of the French Health insurance regime, the EGB (Echantillon généraliste des bénéficiaires). A study is carried out on the complete SNIIRAM database (DCIR) to assess the prediction method performances compared to other multichannel sequence analysis.

Keywords: administrative data, Semi-supervised, multichannel, sequence analysis

### **Determining influential factors in spatio-temporal models**

**Rebecca Nalule Muhumuza<sup>1,2,3</sup>, Olha Bodnar<sup>1</sup>, Sergei Silvestrov<sup>1</sup>,  
Joseph Nzabanita<sup>4</sup>, Rebecca Nsubuga<sup>5</sup>**

<sup>1</sup>*Division of Applied Mathematics, Mälardalen University, Västerås, Sweden,*

<sup>2</sup>*Department of Mathematics, Busitema University, Tororo, Uganda, East Africa,*

<sup>3</sup>*Department of Mathematics, Makerere University, Kampala, Uganda,*

<sup>4</sup>*Department of mathematics, CST-University of Rwanda, <sup>5</sup>Uganda Virus  
Research Institute, Entebbe*

In various areas of modern statistical applications such as in Environmetrics, Image Processing, Epidemiology, Biology, Astronomy, Industrial Mathematics, and many others, we encounter challenges of analyzing massive data sets which are spatially observable, often presented as maps, and temporally correlated. The analysis of such data is usually performed with the goal to obtain both the spatial interpolation and the temporal prediction. In both cases, the data-generating process has to be fitted by an appropriate stochastic model which should have two main properties: (i) it should provide a good fit to the true underlying model; (ii) its structure could not be too complicated avoiding considerable estimation error appeared by fitting the model to real data. Consequently, achieving the reasonable trade-off between the model uncertainty and the parameter uncertainty is one of the most difficult questions of modern statistical theory. We deal with this problem in the case of general spatio-temporal models. New approaches are developed to determine the most influential factors to be included into the model. We also discuss the computational aspects in the case of large-dimensional data and apply the theoretical findings to real data.

Keywords: hierarchical model, latent process, statistical inference, variable selection

## **A partial unemployment among youths in India owing to under reporting of age of older cohorts**

**Barun Kumar Mukhopadhyay**

*Population Studies Unit, Indian Statistical Institute, India (Retired)*

A great section of the older cohort of population still remain in government services, particularly faculties either in colleges or universities because of escalation of retirement age by the ministry of human resources, GOI, time to time from 58 years to 60 to 62 and finally at present to 65 years. The very old pattern of reckoning of ages in the long past in India was based on verbatim since in those times people do not bother about their birth registration consequently there were no official birth records. As it is conjectured that while filling the form for Matric (now school final or 10+) examination in the past parents (or guardians) usually asked their sons and daughters to report their lower age than the original one in the prescribed form so as to remain in services for longer periods, to some extent, in future. Unfortunately, the present young generation are not getting those facilities since for the last couple of years birth certificates are essential for the same purpose while their parents or anybody older generation at the same time having had their ages younger than their original ones because of the reason as mentioned above for remaining in service enjoying the full facility of escalation of retirement age. In the present paper an attempt is made how to estimate the error in age distribution of population of older ages (50 years or more) and compare their distribution with some corrected distribution generated through different methods including UN methods, Graduation techniques based on life tables or if necessary using Lagrange formula, Spread multipliers and others. What proportion of the population of 60 or 65+ years over the adjusted one may be the estimates of proportion of population of young generation (say 18-30/35 years) who are expected to lose their government jobs etc. The study was done for the major states in India. In recent time unemployment rates are growing high. Even some international agency's report is worth mentioning. The sources of data for this kind of analysis may be the UN's Demographic Yearbook, Censuses of India, NFHS, CSOs (GOI), Life tables and others as and when

necessary. The study is under process of collecting data, analysis and final result is yet to come.

**Keywords:** older cohort, Matric, UN methods, Demographic Yearbook

### **A Comparative Study for Forecasting Stochastic Volatility Models: EWMA model versus Heston model**

**Jean-Paul Murara<sup>1,2</sup>, Anatoliy Malyarenko<sup>1</sup>, Milica Rancic<sup>1</sup>, Ying Ni<sup>1</sup>, Sergei Silvestrov<sup>1</sup>**

*<sup>1</sup>Division of Applied Mathematics, School of Education, Culture and Communication, Malardalen University, Vasteras, Sweden, <sup>2</sup>College of Science and Technology, School of Science, Departement of Applied Mathematics, University of Rwanda, Kigali, Rwanda*

Using daily exchange rates of various currencies with respect to Rwandan Franc (RWF), we focus our study in searching whether when forecasting volatility there exists a considerable correlation between the Exponential Weighted Moving Average model ( $I=0.94$  and  $I=0.97$ ) and the Heston model. To test how important that correlation is, we apply some test techniques. Our results show statistically the significant level of the correlation in the forecasts done.

**Keywords:** Stochastic Volatility, Forecasting, Foreign Exchange Option, EWMA, Heston Model

### **Forecasting Stochastic Volatility for Exchange Rate**

**Jean-Paul Murara<sup>1,2</sup>, Anatoliy Malyarenko<sup>1</sup>, Milica Rancic<sup>1</sup>, Ying Ni<sup>1</sup>, Sergei Silvestrov<sup>1</sup>**

*<sup>1</sup>Division of Applied Mathematics, School of Education, Culture and Communication, Malardalen University, Vasteras, Sweden, <sup>2</sup>College of Science and Technology, School of Science, Departement of Applied Mathematics, University of Rwanda, Kigali, Rwanda*

In risk management, foreign investors or multinational corporations are highly interested to know how volatile a currency is in order to hedge risk. Using daily exchange rates, in this paper we perform volatility forecasts for three periods: December 2018, September 2018 to December 2018

and June 2018 to December 2018. We perform the forecasts helped by Exponential Weighted Moving Average (EWMA) model. Based on the results, conclusions are given.

**Keywords:** Stochastic Volatility, Forecasting, Foreign Exchange Option, EWMA

## Pricing Options under two-dimensional Black-Scholes Equations by using C-N Scheme

Jean-Paul Murara<sup>1,2</sup>, Anatoliy Malyarenko<sup>1</sup>, Ying Ni<sup>1</sup>, Sergei Silvestrov<sup>1</sup>

<sup>1</sup>*Division of Applied Mathematics, School of Education, Culture and Communication, Malardalen University, Vasteras, Sweden,* <sup>2</sup>*College of Science and Technology, School of Science, Departement of Applied Mathematics, University of Rwanda, Kigali, Rwanda*

In the option pricing process, Black-Scholes (1973) solved a partial differential equation (PDE) and introduced a model to determine the price of options. While dealing with many problems in financial engineering, the application of PDEs is fundamental to explain the changes that occur in the evolved systems. In this paper, we consider the option-pricing problem that involves a two-dimensional Black-Scholes PDE. With some simulations, we solve the equation using Crank-Nicolson scheme, study its stability and comparing examples are also included in the paper.

**Keywords:** Stochastic Volatility, 2D Black-Scholes PDE, Crank-Nicolson Method

## **A Mathematical Model for Cannibalism in a Predator-Prey System with Harvesting**

**Loy Nankinga**

*Department of Mathematics, Makerere University, Uganda, Kampala, Uganda*

In this paper, we study the interactions of two consumers-resource system with harvesting, in which African catfish and Tilapia consume a shared food resource. The African catfish is cannibalistic but also feed on Tilapia and the food resource, whereas Tilapia feeds on the food resource. A system of ordinary differential equations is developed using unstructured population models. The solutions of these equations are studied in connection with harvesting strategies with stability analyses. We will also study the benefits and drawbacks of considering one-species versus two-species farming system in economic terms. The main goal of our findings will be used to inform policy in order to improve fish harvesting strategies and hence increase on the fish biomass production in Uganda.

**Keywords:** Cannibalism, Predation-Prey, harvesting and unstructured population

## **Stochastic Modeling for Weather Derivatives and Application to Insurance**

**Clarinda Nhangumbe<sup>1</sup>, Alex Marrime<sup>1</sup>, Betuel Canhanga<sup>1</sup>, Calisto Guambe<sup>1</sup>**

*<sup>1</sup>Faculty of Sciences, Department of Mathematics and Computer Sciences, Eduardo Mondlane University, Maputo, Mozambique*

Weather derivatives are financial instruments emerging and growing dynamically in the financial market. They were introduced in the last decades in order to help companies to reduce income variability resulted on adverse weather condition. The temperature level, the amount of rainfalls, snowfalls, frost and winds can be observed in certain period of time and in a certain station in order to build an index on which a payoff of a financial instrument can be based. This type of derivatives can be in our days applied in many sectors, among them, in agriculture, sport, leisure, entertainment, tourism, construction. These derivatives can be offered as options, futures, swaps, etc., but in the pricing process one must consider



that the weather factors are not tradable assets and are random processes. In the present paper we will propose a model for weather derivative by considering stochastic variation of the weather (rainfalls) and we suggest a stochastic differential system describing its evolution. Since the rainfalls is non-tradable quantity, we build a contract in an incomplete market and we determine the analytic solution of the pricing problem. We finish the paper by considering some daily situations where one can apply such type of contingent claim as an insurance contract to mitigate the risk evolved in activities that depends on the amount of rainfalls.

**Keywords:** weather derivatives, incomplete markets, actuarial principles

### Goodness-of-fit tests for logistic family via characterization

Yakov Nikitin<sup>1</sup>, Ilya Ragozin<sup>1</sup>

<sup>1</sup>*Department of Mathematics and Mechanics, Universitetskaya nab. 7/9, Saint-Petersburg State University, Russia*

The logistic family of distributions belongs to the class of important families in Probability and Statistics. However, the goodness-of-fit tests for the composite hypothesis on belonging to the logistic family with unknown location parameter against the general alternatives are almost unexplored. We propose two new goodness-of-fit tests, the integral and the Kolmogorov type, based on the recent characterization of logistic family due to Hu and Lin. They are build using the U-empirical measures. We discuss asymptotic properties of new tests such as their limiting distributions and large deviations, and calculate their local Bahadur efficiency against natural alternatives. Conditions of local asymptotic optimality of new tests are also explored.

**Keywords:** logistic distribution, goodness-of-fit, U-statistics, characterization, asymptotic efficiency

## An Algebraic Method for Pricing Financial Contracts in the Post-Crisis Financial Market

Hossein Nohrouzian<sup>1</sup>, Anatoliy Malyarenko<sup>1</sup>, Ying Ni<sup>1</sup>, Christopher Engström<sup>1</sup>

<sup>1</sup>Division of Applied Mathematics, Mälardalen University, Västerås, Sweden

Before the financial crisis of 2007, the forward rate agreement (FRA) contracts could be perfectly replicated by overnight indexed swap (OIS) zero coupon bonds. After the crisis, the simply compounded risk-free OIS forward rate became less than the FRA rate. Using the approach by Cuchiero et al (2016), we construct an arbitrage-free market model, where the forward spread curves for a given finite tenor structure are described as a mild solution to a boundary value problem for a system of infinite-dimensional stochastic differential equations. The constructed financial market is *large*: it contains infinitely many OIS zero coupon bonds and FRA contracts with all possible maturities. To solve the above system, we use an algebraic approach by Bayer and Teichmann (2008) called the *cubature on Wiener space*.

**Keywords:** Forward Rate Agreement, Overnight Index Swap, Large Market, Cubature, Wiener Space

### References:

- [1] Cuchiero, C., Klein, I., Teichmann, J. A new perspective of the fundamental theorem of asset pricing for large financial markets. *Theory Probab. Appl.* **60** (2016), no. 4, 561–579.
- [2] Bayer, C., Teichmann, J. Cubature on Wiener space in infinite dimension. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **464** (2008), no. 2097, 2493–2516.

## On Dimensionality Reduction and Modelling of Pension Expenditures in Europe

Kimon Ntotsis<sup>1</sup>, Marianna Papamichael<sup>2</sup>, Peter Hatzopoulos<sup>1</sup>, Alex Karagrigoriou<sup>1</sup>

<sup>1</sup>University of the Aegean, Samos, Greece, <sup>2</sup>National Actuarial Authority of Greece

The aim of this work is to locate, collect and analyze the factors which either on short-term or on long-term may have an impact on the shaping of the Pension Expenditures for various European countries. By achieving that we are able to model the Pension Expenditures (as percentage of GDP) and make forecasts.

For this purpose, advanced multivariate techniques are applied to a data set of 20 explanatory variables and 20 European countries for the period 2001-2015.

**Keywords:** Pension Expenditures, Modelling and Forecasting, PCA

## An approach to nonparametric curve fitting with censored data

Jesus Orbe<sup>1</sup>, Jorge Virto<sup>2</sup>

<sup>1,2</sup>Department of Econometrics and Statistics, University of the Basque Country UPV/EHU, Bilbao, Spain

In this study we consider the problem of nonparametric curve fitting in the specific context of censored data. We propose an extension of the penalized splines approach using the Kaplan-Meier weights to take into account the effect of censorship and generalized cross-validation techniques to choose the smoothing parameter adapted to the case of censored samples. Using various simulation studies we analyze the effectiveness of the censored penalized splines method proposed and show that the performance is quite satisfactory. Therefore the methodology proposed is a good alternative when the functional form of the covariate is not known in censored regression models.

**Keywords:** Censored data, Kaplan-Meier weights, nonparametric estimation, penalized splines

## Stochastic comparisons of contagion risk measures in portfolios of dependent risks

Patricia Ortega-Jimenez<sup>1</sup>, Miguel A. Sordo<sup>2</sup>, Alfonso Suárez-Llorens<sup>3</sup>

<sup>1,2,3</sup>*Department of Statistics and Operation Research, University of Cádiz, Spain*

A topic of increasing interest in portfolio risk analysis is the evaluation of risk contagion, which refers to judge how the risk behavior of some components spreads to others or even to the whole portfolio. In this framework, we study the consistency of some recently introduced contagion risk measures, including the marginal expected shortfall (MES) and the marginal mean excess (MME), with respect to various stochastic orderings under different dependence assumptions. We illustrate the applicability of the results in the context of parametric families of distributions, by showing how changes in the parameters affect the risk of contagion.

**Keywords:** Marginal mean excess; Multivariate conditional tail expectation; conditional distribution; dependence; stochastic ordering; contagion risk

## Modeling with Hyperbolic Restrictions: The Nigerian Population Dynamics

Oyamakin Samuel Oluwafemi, Osanyintupin Olawale Dele

*Department of Statistics, Faculty of Science, University of Ibadan, Nigeria*

Intercensal estimate is an estimate of population between official census dates with both of the census counts being known. This was observed for three cases using three growth models so as to determine the effectiveness of models in predicting correctly the census figure. Case 1 was the use of the 1963 population census result as the base year and 1991 population census result as the launch year. Case 2 was the use of the 1991 population census result as the base year and 2006 population census result as the launch year and case 3 was the use of the 1963 population census result as the base year and 2006 population census result as the launch year. The Nigeria population census figure for the year 1963, 1991 and 2006 were used for intercensal prediction while nonlinear estimation was applied on the data sourced online from 1955-2016 for

model validation. A modified Hyperbolic Exponential Growth Model (HEGM) was used along with Exponential Growth Model (EGM) in predicting population figures and Mean Square Error (MSE), Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) were used to assess the suitability of the model on population prediction. Different values of shape parameter in the hyperbolic model were assumed to be small, moderate and high with  $\pm 0.1$ ,  $\pm 0.5$  and  $\pm 0.9$  for Case 1, 2, and 3. HEGM gave the best intercensal estimate for the three cases and was preferred based on the AIC, BIC and MSE results with theta stabilized at  $\pm 0.1$ .

**Keywords:** Intercensal, Growth models, Hyperbolic Growth Model, Nigeria Population

## **Model of Lifetable Evolution with Variable Drift and Cointegration**

**Wojciech Otto**

*Department of Economic Sciences, University of Warsaw, Długa str.  
44/50, 00-241 Warsaw, Poland*

The aim of the paper is to present selected issues arising when looking for an adequate model for long-term predictions of lifetables. Experience comes from analysis of Polish national lifetables for the period 1958-2017, and ages 0-94.

At the first stage Lee-Carter-type models have been fitted to the data. For each gender three calendar effects have been extracted, representing evolution of mortality of (partly) separated age groups: the young, the adult, and the old. Additionally, strong autocorrelation of disturbances captures a kind of “local” cohort effects.

The second stage consists in looking for a 6-dimensional time series model, good enough to capture long-term evolution. Most important differences between alternative models concern various specifications of random effects responsible for stochastic changes of the intensity of drift, as well as various assumptions about cointegration of the series. Special attention is paid to techniques of analyzing and illustrating long-term properties of alternative models. Looking for adequate specification is driven by three most important questions:

- to what extent changes of mortality observed within subgroups {young, adult, old}  $\times$  {males, females} are interrelated,

- do the slope of trends vary in time significantly,
- are errors of long-term prediction large

A positive answer is given to all three questions, however, the strongest is empirical evidence supporting the answer to the first one. Empirical evidence supporting positive answers to next two questions is weaker. The best model (in terms of fit and some desired properties of long-term predictions) supports positive answers, and renders long-term prediction errors much greater than traditional simple models. However, alternative models rendering much smaller uncertainty are not as much worse in terms of fit within the sample.

Some methodological results are of general interest. Other results are less general, as they are due to specific properties of national lifetables, long observation period, wide range of ages, and specific country.

**Keywords:** Lee-Carter model, Haberman-Renshaw model, stochastically varying parameters, Kalman Filter, cointegration

## On the evaluation of 'Self-perceived Age' for Europeans and Americans

Apostolos Papachristos<sup>1</sup>, Georgia Verropoulou<sup>2</sup>

<sup>1</sup>Department of Statistics and Insurance Science, University of Piraeus, Greece,

<sup>2</sup>Department of Statistics and Insurance Science, University of Piraeus, Greece

The aims of the study are to estimate 'Self-perceived age' by reference to life tables and to evaluate its validity in comparison with actual mortality patterns. We use data from the 6th Wave of the Survey of Health, Ageing and Retirement in Europe (RAND SHARE), the 12th Wave of Health and Retirement Study (RAND HRS) and life tables from the Human Mortality Database (HMD). For the statistical analysis we employ regression models. Our results indicate that health status and frequency of physical activities imply similar patterns of 'Self-perceived age' and actual mortality patterns. Individuals with better health tend to have younger 'Self-perceived age' and lower actual mortality. However, the impact of memory and cognitive function differentiates between Europeans and Americans. 'Self-perceived age' is expressed in years, is linked to a population life table and it could be used to detect early changes in future life expectancy.

**Keywords:** Self-perceived age, Subjective survival probabilities, HRS, SHARE, HMD, welfare states

This work has been partly supported by the University of Piraeus Research Center.

## **Analysing the risk of bankruptcy of firms: survival analysis, competing risks and multistate models**

**Francesca Pierri<sup>1</sup>, Chrys Caroni<sup>2</sup>**

*<sup>1</sup>University of Perugia, Perugia, Italy, <sup>2</sup>National Technical University of Athens, Greece*

The interest and research activity in Credit Scoring techniques and its application have increased during the last decade following the implementation of Basel II agreements, concurrently with the severe economic crisis that has affected Europe and the world. Quantitative methods have been applied to risk analysis since 1966 with Beaver, starting with Multivariate Discriminant Analysis and subsequently Logistic Regression. Firms or customers are assigned to two different groups, bad or good, on the basis of the probability of failure/success predicted by the model. Narain in 1992 first had the insight to apply survival analysis to credit risk modelling in order to study the time spent by a subject in the healthy group and from then onwards several economic studies have taken into account this methodology, originally applied in the fields of medicine and engineering. After a brief review of quantitative methods applied in Credit Scoring, this paper focuses on analyzing different causes of failure. We will present the use of competing risk methodology from survival analysis describing two different approaches. Furthermore, we will go on to take into account the occurrence of a second event in addition to the first and we will study transition and survival probabilities applying a multistate model. A large data set is used to demonstrate the application of these methodologies to predicting the bankruptcy of Small and Medium Enterprises.

**Keywords:** Bankruptcy, Logistic Regression, Survival Analysis, Competing Risks, Multistate Model

## Ensemble Methods for preference structures, with relevance to the rankings' positions

**Antonella Plaia, Simona Buscemi, and Mariangela Sciandra**

*Department of Economics, Business and Statistics, viale delle Scienze, ed. 13,  
Palermo, Italy*

In the framework of preference rankings, one main research interest is focused on identifying the subject profiles having similar preferences. This implies to identify those predictors able to explain the observed preference structure. Moreover, in many real situations the true research could be interested only into the top or the bottom of the observed rankings.

This paper aims to reply to the combination of these two aspects.

As concerns the first one, decision trees could be a good solution, even if a single tree could result unstable and not very accurate.

With reference to the second point, in the framework of decision trees, the use of a position-weighted rank correlation coefficient as impurity function and as a tool for detecting the consensus ranking, ensures more homogeneity among units, if the attention is paid only on the first or last rankings' positions. In order to improve stability and accuracy, ensemble methods have been called the most influential development in Data Mining and Machine Learning in the past decade. They combine multiple models into one usually more accurate than the best of its components.

In order to identify correctly similar groups of units (in terms of their preferences) and to take into account the relevance given to the rankings' position, two ensemble methods are suitably modified and proposed in this work: boosting and bagging. These procedures are well known in literature, except when the objective variable is a ranking and especially if not all the rankings' positions are equally relevant.

This work shows the theoretical and practical aspects of the proposed ensemble methods through examples both simulated and real ones.

**Keywords:** Bagging, Boosting, Ranking data, position weighted rank correlation coefficient.



## Estimating age-demographic trends based on Renyi entropy

Vasile Preda, Irina Bancescu

*National Institute of Economic Research "Costin C. Kiritescu", Bucharest, Romania*

Entropic measures play an essential role in various fields such as economics, informatics, engineering, medicine and physics. Most known entropy is Shannon entropy introduced in 1948. Since then many new entropic measures have been introduced, such as Tsallis, Varma, relative and weighted entropies. A generalization of Shannon entropy is Renyi entropy which we will use to estimate the demographic trends of Romania's population. According to many studies, Romania has a large mobility, within and outside the country. Predicting demographic trends is a crucial and open research topic. In 2015, Zhao G.S et al. [1] proposed an entropy-based method for demographic research which involved three stages. We extend this method by introducing Renyi entropy into the equation. This method takes into account the age-dependent structure of the population.

**Keywords:** demographic trends; entropic measures; age-dependent structure

### References:

[1] Zhao, Guang-She, Yi Xu, Guoqi Li, and Zhao-Xu Yang. "An entropy-based method for estimating demographic trends." In *Evolving and Adaptive Intelligent Systems (EAIS)*, 2015 IEEE International Conference on, pp. 1-8. IEEE, 2015.

## Latest frontiers in grouped-ordinal data dependence analysis

Emanuela Raffinetti<sup>1</sup>, Fabio Aimar<sup>2</sup>

*<sup>1</sup>Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, Milano, Italy, <sup>2</sup>School of Management and Economics, C.so Unione Sovietica 218 Bis, Turin, Italy - ASL CN1, Cuneo, Italy*

The bivariate dependence analysis is strongly supported in literature by a wide set of measures, including the Pearson's  $r$ , the Kendall's  $\tau_b$  and the Spearman's  $r_s$  correlation coefficients among others. Currently, we are

assisting to an explosion in the availability of ordinal data due to widespread attitudinal surveys. In many cases, survey scales are also built on responses that are observed to belong to certain groups on a continuous scale (grouped variable). Given  $h$  groups, the measurement problem may be addressed by encoding each group through a label (from 1 to  $h$ ) and, subsequently, by assigning rank one to all the units included in the first ordered group and finally rank  $h$  to those included in the  $h$ -th ordered group. In such a way, the assessment of the direct or inverse dependence relationship may be carried out through Spearman's  $r_s$  (e.g. Spearman [3]) or Kendall's  $\tau_b$  (e.g. Kendall [2]) coefficients which are based on the correlation between the ranks of two variables and on the pairs of concordant and discordant values of two variables, respectively. This results in neglecting the original continuous nature of the grouped variable, since the information from the grouped variable has to be reduced to its ordinal information, too. A crucial issue is then related to dependence relationship studies when one variable is ordinal and the other variable is grouped. The "Monotonic Dependence Coefficient" (MDC), recently proposed by Ferrari and Raffinetti [1], is here re-formalized for the case of grouped and ordinal variables. Through a Monte Carlo simulation study, some basic hints about the new MDC coefficient performance in specific scenarios are given even in comparison with Spearman's and Kendall's coefficients. The contribution ends with an application to drug-expenditure data incurred by the Italian system for public health assistance, whose aim is to illustrate the role of age differences in the allocation of drug expenditure both by considering overall patients and single sub-groups, differing in terms of gender.

**Keywords:** dependence analysis, grouped data, ordinal data, Monte Carlo simulation

**References:**

1. PA Ferrari, E. Raffinetti. A different approach to Dependence Analysis. *Multivariate Behavioral Research*, 50, 2, 248-264, 2015.
2. K. Kendall. New Measure of Rank Correlation. *Biometrika*, 30, 1/2, 81-89, 1938.
3. C. Spearman. The proof and measurement of correlation between two things. *American Journal of Psychology*, 15, 72-101, 1904.

## Hybrid multiple imputation for incomplete household surveys

**Humera Razzak<sup>1</sup>, Christian Heumann<sup>2</sup>***<sup>1,2</sup>Institut für Statistik Ludwig-Maximilians-Universität München, Deutschland*

Large scale surveys e.g. (MICS) often contains significant amount of item non response. Fully conditional specification (FCS) multiple imputation (MI) approach can fail to impute such data due to compatibility and complex dependencies among categorical variables whereas joint modeling (JM) MI approach is limited to only categorical variables and requires transformations (or other tricks) for continuous variables. We propose a simple and easy to implement hybrid MI approach which combines both existing MI approaches. Purposed hybrid technique uses the the information available on categorical variables to impute continuous variables and vice versa. Hybrid MI method performs better as compared to the existing MI methods in simulation studies. Results in simulation studies are supported by a household data example from MICS 2014.

**Keywords:** Survey data; Multiple Imputation; Household data; Hybrid; MIC

## **Weak Signals in High-dimensional Poisson Regression Models**

**Orawan Reangsephet<sup>1</sup>, Supranee Lisawadi<sup>2</sup>, S. Ejaz Ahmed<sup>3</sup>**

*<sup>1</sup>Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani, Thailand, <sup>2</sup>Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani, Thailand, <sup>3</sup>Department of Mathematics and Statistics, Faculty of Mathematics and Science, Brock University, St. Catharines, Ontario, Canada*

In this work, we addressed parameter estimation and prediction for the sparse Poisson regression model under high-dimensional regimes in which the number of predictors/features ( $p$ ) exceeds the sample size ( $n$ ). Generally, the dimensionality reduction via the penalized maximum likelihood approaches is a critical stage, before making post-selection parameter estimation based on the resulting model from the dimensionality reduction stage via maximum likelihood (ML). The key point is that the use of different approaches results in a different subset of selected predictors, usually of unknown correctness. This may produce either overfitted or underfitted models, making post-selection ML estimators based on these models inefficient. Hence, we proposed the post-selection estimators based on linear shrinkage, pretest, and Stein-

type shrinkage strategies to improve the performance of classical ML estimators based on the models obtained from the dimensionality reduction stage. Through Monte Carlo simulations, the results demonstrated that the proposed estimators were shown to be significantly more efficient than the classical ML estimators, regardless of the correctness in the dimensionality reduction stage.

**Keywords:** Poisson regression, Monte Carlo simulations, Penalized maximum likelihood, Linear shrinkage, Pretest, Stein-type shrinkage

## Some New Results in Bandit and Related Problems

**Sheldon Mark Ross**

*University of Southern California, USC Viterbi School of Engineering*

*Daniel J Epstein Department of Industrial and Systems Engineering*

We discuss some recent results in bandit related models. We discuss a strategy that combines fiducial probability with Thompson sampling. Applications to contextual bandits and to one where it is known that the parameter of interest is a unimodal function of the bandit arm used will be given. We also discuss dueling bandits, where each arm has an unknown value, and where at each stage two arms are chosen to play a game, with the team with value  $v$  beating one with value  $w$  with probability  $v/(v+w)$ . With the objective being to choose successive pairs so as maximize the number of times that the two highest value arms are chosen, we present a Thompson sampling type procedure for making choices. We also present a method for estimating the values, a problem which has applications in a variety of sports.

## Response-Adaptive Randomization: Optimizing Clinical Trials for Ethics and Efficiency

**William F. Rosenberger**

*Department of Statistics, George Mason University, Fairfax, USA*

We review response-adaptive and covariate-adaptive randomization procedures and then describe a hybrid design called covariate-adjusted response-adaptive (CARA) randomization. The goal of CARA randomization is to assign more participants in the clinical trial the treatment that is best for them, according to their covariate profile. We review three types of designs: (1) designs based on optimal allocation targets; (2) designs based on the Gittins' index; and (3) designs based on urn models. CARA is such a new topic that very little is known about the properties of these designs. We discuss what is known, and the fertile ground for open problems that such designs present. We also describe how these designs might be applied to precision medicine and enrichment designs.

**Keywords:** Response-adaptive Randomization, Covariate-adaptive Randomization, Clinical Trial Design

## Robust Bayesian analysis using multivariate classes of priors distributions

**F. Ruggeri<sup>1</sup>, M. Sánchez-Sánchez<sup>2</sup>, M.A. Sordo<sup>3</sup>, A. Suárez-Llorens<sup>4</sup>**

<sup>1</sup>*IMATI (Istituto di Matematica Applicata e Tecnologie Informatiche "Enrico Magenes") – CNR (Consiglio Nazionale delle Ricerche), Milano, Italy,*

<sup>2,3,4</sup>*Department of Statistics and Operational Research, University of Cadiz, Spain*

This talk generalizes to the multivariate setting some ideas recently developed in Sánchez-Sánchez et al. (2018) in the framework of univariate robust Bayesian analysis. By weighting a particular prior belief, we introduce a class of multivariate prior distributions that fulfills some desirable properties. Then, we study the propagation of uncertainty from this class to the associated class of posterior distributions. Uncertainty of these classes is evaluated by different metrics, such as the Hellinger metrics and the Kullback-Leibler divergence. An application with real data involving failure times in reliability systems is given.

**Keywords:** Robust Bayesian Analysis, prior class, stochastic orders, weight functions, reliability

**References:**

1. Sánchez-Sánchez, M., Sordo, M.A., Suárez-Llorens, A. and Gómez-Déniz, E. Deriving robust Bayesian premiums under bands of prior distributions with applications. *Astin Bulletin*, doi: <https://doi.org/10.1077/asb>. 2018.36 , 2018.

## Heterogeneity of chronic pathology burden among elderly

**Tamara Sabgayda<sup>1</sup>, Anna Edeleva<sup>2</sup>, Victorya Semyonova<sup>1</sup>**

*<sup>1</sup>Department of analysis of health statistics, Federal Research Institute for Health Organization and Informatics of Ministry of Health of Russian Federation, Moscow, Russia, <sup>2</sup>Lobachevsky State University of Nizhni Novgorod, Russia*

Russian elderly suffers from a burden of different chronic diseases. At that, the system of collection and processing of morbidity data doesn't provide insight of prevalence of primary and comorbid conditions and complications which disturbs health profile and elaboration of adequate healthcare measures. **Materials and methods.** The study was based on doorbell survey of households including elderly (60 years and older for men and 55 and older for women), sample of 14749 urban and 7808 rural residents. The elderly was examined by medical teams using medical documentation on determined diagnosis, asked about self-estimation of health and behavioural risk factors. The study subject is Nizhny Novgorod region where life expectancy level is close to Russia's average.

**Results.** Profiles of comorbid conditions and complications were constructed for cardio-vascular diseases, neoplasms and diabetes mellitus serving as primary diagnosis. Absence of common patterns was shown for nosology structure of comorbid conditions and its dependence from primary diagnosis, sex and age of an elderly person, his place of residence and presence of behavioural risk factors. Not just list but even a number of comorbid conditions depends upon primary diagnosis reaching its maximum in men with cardio-vascular diseases ( $3.58 \pm 0.2$ ), and in women with diabetes mellitus ( $2.58 \pm 0.15$ ). Summary registration of primary diseases and comorbid conditions bring to leading positions such

diseases as diseases of musculoskeletal, digestive, endocrine and genitourinary systems apart from cardio-vascular diseases.

**Conclusions.** Complementation of medical statistics with data of sampling studies modifies health profiles especially in the elderly. Primary disease is a factor in clustering the burden of chronic diseases.

**Keywords:** health of elderly, burden of chronic diseases, primary disease, comorbid condition, structure of chronic diseases

## Statistical Analysis of Data from Experiments Subject to Restricted Randomisation

**Aljeddani Sadiah**

*Umm Al Qura University, Mathematical Science school: Statistic  
Saudi Arabia*

The selection of the best subset of variables, which will have a strong effect on an outcome of interest, is fundamental when avoiding overfitting in statistical modelling. However, when there are many variables, it is computationally difficult to end this best subset. The difficulties of variable selection would be more complex when designs are with restricted randomisation. This work aims to fill the gap of variable selection and model estimation for data from experiments subject to restricted randomisation by developing new methods for variable selection and model estimation using frequentist analysis and Bayesian analysis for experiments subject to restricted randomisation. Frequentist and Bayesian analysis methods are used to carry out a comparative study with respect to their performance in variable selection and model estimation. As a representative of frequentist analysis, the Penalised Generalised Least Square (PGLS) estimator is used in which a single shrinkage parameter is applied to all regression effects. Furthermore, as two different strata in split-plot design are existed, the PGLS approach is extended to perform variable selection and model estimation simultaneously in the context of split-plot design. The Penalised Generalised Least Squares for Split-Plot Design estimator (PGLS-SPD) is utilized, in which two shrinkage parameters are applied, one for the subplot effects and the other for the whole-plot effects. As a representative of Bayesian analysis, the Stochastic Search Variable Selection (SSVS) technique is used. This performs variable selection and model estimation simultaneously where the variance of all active factors will be sampled

from one posterior distribution. As two different strata in split-plot design are existed, the SSVS approach to perform Bayesian variable selection is extended for the analysis of data from restricted randomised experiments by introducing the Stochastic Search Variable Selection for Split-Plot Design (SSVS-SPD) in which the variances of the active subplot and whole-plot factors are sampled from two different posterior distributions. The usefulness of frequentist and Bayesian approaches are demonstrated using two practical examples, and their properties are studied in simulation studies. The result of the comparative study of frequentist analysis and Bayesian analysis supports the utilization of SSVS-SPD method for the statistical analysis of data from experiments subject to restricted randomisation.

**Keywords:** Split-plot design, Frequentist analysis, Bayesian analysis.

### **Threshold Regression Model with Applications to the Adherence of HIV Treatment**

**Takumi Saegusa<sup>1</sup>, Ying Qing Chen<sup>2</sup>, Mei-Ling Ting Lee<sup>3</sup>**

<sup>1</sup> *University of Maryland, MD, USA*

<sup>2</sup> *University of Washington, WA, USA*

<sup>3</sup> *University of Maryland, College Park, MD, USA*

Cox regression methods are well-known for time-to-event analysis. It has, however, a strong proportional hazards assumption that might not always hold in applications. In many medical contexts, a disease progresses until an event (such as onset of disease or death) is triggered when the health level first reaches a failure threshold. I'll present a Threshold Regression (TR) model for the health process that requires few assumptions and, hence, is quite general in its potential application. A case example on the adherence of antiretroviral treatment for HIV will be presented to demonstrate its practical use.

Antiretroviral pre-exposure prophylaxis (PrEP) and treatment as prevention (TasP) have been shown to be promising in preventing sexual transmission of human immunodeficiency virus (HIV). An effective intervention depends highly on the adherence to antiretroviral treatment (ART). Using threshold regression results, we found significant factors associated with adherence and estimated the mean adherence time (i.e., time-to-first-non-adherence). Our findings can serve as a basis for



planning the intervention for adherence for successful implementation of HIV prevention programs.

## Health loss among the late pre-retirement and early retirement population

**Victorya Semyonova<sup>1,2</sup>, Tamara Sabgayda<sup>1</sup>**

<sup>1</sup>*Department of analysis of health statistics, Federal Research Institute for Health Organization and Informatics of Ministry of Health of Russian Federation, Moscow, Russia,* <sup>2</sup>*The Institution of the Russian Academy of Sciences the Institute of Socio-Political Research RAS (IRAS ISPR), Moscow, Russia*

The Russian age structure that underwent changed in the 2000s made it necessary to increase the retirement age threshold by 5 years in both males and females (from 60 to 65 and from 55 to 60 years, respectively); the increased life expectancy by 7.4 and 4.8 years up to 66.5 and 77.1 years made this increase substantialized both within the demographic and economic contexts. However, in the light of the upcoming increase in the retirement age threshold, the question seems only natural: how does the fact of retirement affect health of the Russian population? With age, the picture may change in line with the following two scenarios: according to the first one, retirement is characterized by dramatic changes in health at the early retirement age, according to the second one - these changes are evolutionary in nature. To test these hypotheses, the authors have analyzed changes in the mortality 10 years prior to retirement (among males aged 50-54 and 55-59 and females aged 45-49 and 50-54) and over the first 10 years on pension (among males aged 60-64 and 65-69 and females aged 55-59 and 60-64, respectively). In 2016, mortality among males aged 60-64 was 1.5 fold higher compared to males aged 55-59 (vs. 41.7% increase among males aged 55-59 compared to males aged 50-54). In the female population mortality among the age group of 55-59 was 47% higher compared to the age group of 50-54 against the background of a 35.3% increase in mortality among females aged 50-54 compared to females aged 45-49.

**Conclusion.** Within the present context of health, the mere fact of reaching the retirement age threshold does not lead to any dramatic consequences - there is a certain acceleration of age-adjusted mortality rates against the background of evolutionary changes in the structure of causes of death.

**Keywords:** retirement age, age-adjusted mortality rates, mortality structure, leading causes of death

## Limiting Form for the Ergodic Distribution of a Semi-Markovian Random Walk with General Interference of Chance

Ozlem Ardic Sevinc<sup>1,2</sup>, Tahir Khaniyev<sup>1</sup>

<sup>1</sup>Department of Industrial Engineering, TOBB University of Economics and Technology, Ankara, Turkey, <sup>2</sup>Department of Statistics, Central Bank of the Republic of Turkey, Ankara, Turkey

In this study, a semi-Markovian random walk ( $X(t)$ ) with general interference of chance is constructed and investigated. During this study, asymptotic method is used as main mathematical tool. The key point of this study is the assumption that the discrete interference of chance has a general form. Under some conditions, it is proved that the process  $X(t)$  is ergodic and the exact form of the ergodic distribution of the process  $X(t)$  is obtained. Next, it is shown that the ergodic distribution ( $Q_X(\lambda x)$ ) of the process  $X(t)$  weakly converges to the limiting distribution  $R(x)$ :

$$Q_X(\lambda x) \equiv \lim_{t \rightarrow \infty} P\{X(t) \leq \lambda x\} \xrightarrow{\lambda \rightarrow \infty} R(x) \equiv \frac{1}{E(\zeta_1)} \int_0^x (1 - \pi_1(t)) dt$$

Here, the random variable  $\zeta_1$  expresses the discrete interference of chance and  $\pi_1(t) \equiv P\{\zeta_1 \leq t\}$ .

**Keywords:** Semi-Markovian Random Walk, Discrete Interference of Chance, Ergodic Distribution, Weak Convergence, Limit Distribution

## Correlation and the time interval over which the variables are measured – a non-parametric approach

Amit Shelef<sup>1</sup>, Edna Schechtman<sup>2</sup>

<sup>1</sup> Department of Industrial Management, Sapir Academic College, D.N. Hof Ashkelon 79165, Israel

<sup>2</sup> Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 84105, Israel

It is known that when one (or both) variable is multiplicative, the choice of differencing intervals ( $n$ ) (for example, differencing interval of  $n=7$  means a weekly datum which is the product of seven daily data) affects the Pearson correlation coefficient ( $\rho$ ) between variables (often asset returns) and that  $\rho$  converges to zero as  $n$  increases. This fact can cause the resulting correlation to be arbitrary, hence unreliable. We suggest using Spearman correlation ( $r$ ) and prove that as  $n$  increases Spearman correlation tends to a limit which only depends on Pearson correlation based on the original data (i.e., the value for a single period). In addition, we show, via simulation, that the relative variability (CV) of the estimator of  $\rho$  increases with  $n$  and that  $r$  does not share this disadvantage. Therefore, we suggest using Spearman when one (or both) variable is multiplicative.

**Keywords:** Multivariate statistics; Differencing interval; Pearson correlation; Spearman correlation.

## Information Networks and Perturbed Markov Chains with Damping Components

Dmitrii Silvestrov<sup>1</sup>, Benard Abola<sup>2</sup>, Pitos Seleka Biganda<sup>2,3</sup>, Sergei Silvestrov<sup>2</sup>, Christopher Engstrom<sup>2</sup>, John Mango Magero<sup>4</sup>, Godwin A. Kakuba<sup>4</sup>

<sup>1</sup>Department of Mathematics, Stockholm University, Stockholm, Sweden,

<sup>2</sup>Division of Applied Mathematics, School of Education, Culture and Communication (UKK), Mälardalen University, Sweden, <sup>3</sup>Department of Mathematics, College of Natural and Applied Sciences, University of Dar es Salaam, Dar es Salaam, Tanzania, <sup>4</sup>Department of Mathematics, Makerere University, Kampala, Uganda

Markov chains with damping components are popular models used for analysis of information networks. New results on asymptotic expansions

for stationary distributions and coupling estimates for the rate of convergence in ergodic theorems for regularly and singularly perturbed Markov chains with damping components are presented as well as results of related numerical experiments.

**Keywords:** Information network, Perturbation, Markov chain, Asymptotic expansion, Coupling

## **A Novel Approach for Predicting Quality Sleep Efficiency from Wearable Device's Data**

**Mayank Singh, Viranjay M. Srivastava**

*University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa*

In this information age, the intervention of new technologies is creating problems in various industries like healthcare. Artificial intelligence, machine learning, and automation have the highest impact on healthcare. Around 86% of healthcare provider organization, technology vendors and related companies of healthcare are using artificial intelligence [1]. By 2020, the overall expected expenditure of these organizations will be approximately \$54 million on projects related to artificial intelligence. Artificial intelligence brings a paradigm shift to healthcare by increasing of healthcare data availability and robust raid analytics [2]. With the increased pace of daily living, sleep has become essential to academic and workplace performance. Sleep deprivation can result in catastrophic events for those in professions that require high accuracy and safety levels. Studies also confirm that lack of sleep worsens a variety of health problems—from obesity, diabetes, and sleep apnea to Alzheimer's and cancer.[3,4] Systematic sleep studies have become a high priority to overcome the health issues. Various healthcare applications are available nowadays to help the cancer patients for sleep coaches or healthy person to overcome the sleep disorders. Technologies specially wearable's, provide a crucial role in the development of these applications and analysis of sleep health [5]. Wearable devices can help in capturing and analyzing the quality sleep duration. The predictive methodologies can support the medical practitioners and patients after analyzing these data for behavioral health decision that can lead to better sleep and improved health [6]. The present analysis of sleep time duration and its quality are insufficient and unable to ultimately use the wearables' data for health monitoring and analysis [7, 8]. To overcome this problem, we have

explored an innovative approach to predict the quality sleep duration from wearable's physical activity data. We have combined the deep learning with an algorithm which automatically recognizes the human activity. The series of experiments were conducted to compare our approach with the existing method statistically. Our approach showed remarkable improvement in the statistical predictive region. The results are evidence that our approach can significantly enhance applications that assist patients and medical practitioners in making critical behavioral health decisions.

**Keywords:** Sleep Efficiency, Wearable Device Data, Predicting Technique

## **Efficient Method for Lighting and Blind Control in Smart Homes to Save Energy Consumption**

**Mayank Singh, Viranjay M. Srivastava**

*University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa*

In our time the biggest challenges are growing shortage of resources like energy and climate change. Energy is a most demanding resource for many countries around the world. Approximately 50% of consumed energy is imported which is expected to reach 75% by 2030. To overcome this problem, sustainable and efficient energy usage is the most urgent necessity. Building technologies are the largest energy consumer for lighting, heating and cooling which is approximately 40% to total consumed energy in a nation. There is a lot of scope for efficient energy optimization. For energy optimization, we can use the intelligent building with controllers for lighting, ventilation, air conditioning, and heating in networked rooms. For the optimization of effective energy in the building, several approaches and concepts are probable. In this context, intelligent building technologies provide effective cost-benefit in terms of saving in energy consumption. In this paper, we have proposed efficient control approaches for lighting and blinding based on occupancy sensing, user adaptive control, daylight harvesting, and light level tuning and automated motorized shades. We have proposed a closed loop integrated control of lighting and blinds with certain constraints. With the simulated and experimental result, we justify that our proposed method provides substantial energy efficiency in variety of condition with constraints.

**Keywords:** Energy efficient method, Light control system, Motorized blind control system, Intelligent building

## **Distributional Properties of the Percentage Change of Discrete Valued Stochastic Processes**

**George-Jason Siouris and Alex Karagrigoriou**

*Lab of Statistics and Data Analysis, Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Karlovasi, 83200 Samos, Greece*

After extensive investigation on the statistical properties of financial returns, a discrete nature has surfaced when low price effect is present. In order to model the discrete nature of the returns the discretization of the tail density function is applied. This is a rather logical approach, since the nature of returns is discrete, as the market always operates on a specific accuracy. As a result of this discretization process, it is now possible to improve the expected percentage shortfall estimations.

This discrete nature seems to be useful in a number of scientific fields, hence it is generalized in the Percentage Change of Discrete Valued Stochastic Processes. The exotic behaviour it exhibits, as well as the new possibilities it provides are presented in this work.

**Keywords:** Percentage Change, Discrete Valued Stochastic Processes

## **The Weibull model and its relationship to the Healthy years lost in a human population**

**Christos H Skiadas<sup>1</sup>, Charilaos Skiadas<sup>2</sup> and Konstantinos N. Zafeiris<sup>3</sup>**

*1 ManLab, Technical University of Crete, Chania, Crete, Greece*

*2 Department of Mathematics and Computer Science, Hanover College, Indiana, USA*

*3Democritus University of Thrace, Komotini, Thrace, Greece*

The Weibull distribution is a continuous probability distribution which originally served as a model for material breaking strength. Later, it was found that it can be applied on a variety of data from different sources like demography, biology, economics etc. In this paper we calculate the shape parameter of the Weibull model based on life table data, and afterwards

we present a general model of survival-mortality in which we estimate a parameter related to the Healthy Life Years Lost (HLYL). It was found that the two estimated parameters were in accordance to each other and thus they could serve the estimation of the health of a population. This is because the estimations HLYL of the World Health Organization are very close to the two parameters, thus the validity of the proposed method is unquestionable. Secondly, it was found that the shape parameter of the Weibull model is a specific case of the survival-mortality model that we have developed.

**Keywords:** Weibull distribution, survival-mortality analysis, Healthy Life Years Lost (HLYL)

## Branching Processes as Models of Epidemics with Vaccination Control

**Maroussia Slavtchova-Bojkova**

*Sofia University "St. Kliment Ohridski", Sofia, Bulgaria*

One of the relevant application fields of the beautiful theory of branching processes, among many others in biology, is the epidemics modelling. Recently, a framework for analyzing time-dependent vaccination policies for epidemics, which are modelled by a Crump–Mode–Jagers branching process, have been developed (see Ball et al. (2014)). Stochastic monotonicity and continuity results for a wide class of functions (e.g., extinction time and total number of births over all time) defined on such a branching process are proved, leading to optimal vaccination schemes to control corresponding functions of epidemic outbreaks. We are developing a new model and studying its basic properties, obtaining its Sellke construction (Sellke, (1983)). We are considering a more general framework given by the general SIR (susceptible-infective-removed) epidemic model, as well as some of its generalizations (for example, the SEIR -susceptible-exposed-infective-recovered -epidemic model) or extensions (for example, the SIS -susceptible-infective-susceptible-epidemic model). Finally, we would like to point out, that this reveals the fundamental role of the theory of branching processes in human practice, which would not be possible without the ending part of the general model of branching processes. Acknowledgements. The research is supported by the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria, grant No KP-6-H22/3 and partially supported by and

Ministerio de Economía y Competitividad, and the FEDER through the Plan Nacional de Investigación Científica Desarrollo e Innovación Tecnológica, grant MTM2015-70522-P, Spain.

**Keywords:** General branching processes, SIR epidemic model, Sellke construction, vaccination policies

**References:**

1. Ball, F., González, M., Martínez, R. and Slavtchova–Bojkova, M. (2014). Stochastic monotonicity and continuity properties of functions defined on Crump–Mode–Jagers branching processes, with application to vaccination in epidemic modelling. *Bernoulli*, 20(4), 2076–2101.
2. Sellke, T. (1983) On the asymptotic distribution of the size of a stochastic epidemic. *J. Appl. Prob* 20 (20), 390–394.

## Applied Meta-Analysis in Two-Class Overbooking Model

**Murati Somboon<sup>1</sup>**

*<sup>1</sup>Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand*

In airline industry, the optimal overbooking limit is the key of success in airline revenue management. A two-class overbooking model that combines two of the most important airline revenue management, namely overbooking and seat inventory control for passenger airline is possible to find a closed-form for optimal booking limit and the optimal overbooking limit simultaneously. The optimal booking/overbooking limit was calculated by using the mean demand estimation for class-1 and class-2. In general, mean demand of two classes was calculated by time series method. In this study, a meta-analysis is applied in order to improve the performance of a two-class overbooking model by estimated mean demand of two classes. Meta-analysis is the statistical procedure for combining data from multiple studies. In this case, the data was divided by the update booking limit point to multiple studies. A numerical study was set to evaluate the performance of the two-class overbooking model that applied meta-analysis against an exponential smoothing method. A two-class overbooking model is more outperformed when meta-analysis is applied.

**Keywords:** Overbooking, Meta-Analysis, Static Model, Stochastic Model, Revenue Management, Airline Passenger



## Interpolation with stochastic local iterated function systems

**Soós Anna<sup>1</sup>, Somogyi Ildikó**

*<sup>1</sup>Department of Mathematics and Computer Science of the Hungarian Line,  
Babes Bolyai University, Cluj-Napoca, Romania*

The methods of real data interpolation can be generalized with fractal interpolation. These fractal interpolation functions can be constructed with the so-called iterated function systems. Local iterated function systems are important generalization of the classical iterated function systems. In order to obtain new approximation methods this methods can be combine with the classical interpolation methods. In this paper we focus on the study of the stochastic local fractal interpolation function in the case when the vertical scaling parameter is a random variable.

**Keywords:** fractal functions, attractors, interpolation

### **References:**

1. M.F. Barnsley. Fractal functions and interpolation, Constructive Approximation, vol. 2, 303-329, (1986)
2. M. F. Barnsley, M. Hegeland and P. Massopust, Numerics and Fractals, <https://arxiv.org/abs/1309.0972>, 2014
3. M.F. Barnsley, Fractals Everywhere, Academic Press, (1993).
4. A.K.B.Chand, G.P.Kapoor, Generalized cubic spline interpolation function, SIAMJ.Numer. Anal 44(2), 655-676 (2006).
5. J.E.Hutchinson, Fractals and Self Similarity, Indiana University Mathematics Journal, 30, no.5, 713-747, (1981).
6. I. Somogyi, A. Soós, Stochastic fractal interpolation with variable parameter, International Scentic Journal, Journal of Mathematics, 28-33, (2015).
7. I. Somogyi, A. Soós, Interpolation using Local Iterated Function Systems, International Conference of Numerical Analysis and Approximation Methods, ICNAAM2017, 25-30 Sept. Thessaloniki, Greece, (2017)
8. H.Y. Wang, J.S. Yu, Fractal interpolation functions with variable parameters and their analytical properties, J, Approx Theory 175, 1-18, (2013).

## **Acoustic evaluation of the rise-age in children's acquisition of speech sounds**

**D. A. Sotiropoulos**

*Technical University of Crete, Chania, Crete, Greece*

It is known that typically developing children completely acquire sounds in running speech by about the age of five years in most languages. However, there are children whose speech acquisition is delayed independently of when they start to talk. An area of interest, little researched, is the determination of the length of time that it takes a child with typical or atypical speech development to repeatedly correctly produce speech sounds from correctly producing speech sounds only occasionally; this length of time is hereby called rise-age. This is an important area in child speech research because it will provide a guide as to how long to expect progress to last in typical acquisition of speech sounds. This will help define speech delay so that possible intervention can be sought. There are two questions to be answered in determining rise-age. One is how to define occasional as well as repeated production and the other is to define correctness. Correctness will be determined by comparing the acoustic characteristics of the spectrograms of child produced speech sounds to the acoustic characteristics of adult speech sound spectrograms. A speech sound will be considered correct if there is a 90% spectrogram match between child and adult. In turn, occasionally correct will be the sounds that only 15% or less of the times they are produced are correct, while repeatedly correct will be the sounds that 90% or more of the times they are produced are correct. These definitions have been applied to determine the rise age of a child's speech sounds in two languages, Greek and English. The data comprised speech sounds in the child's digitally recorded running speech during conversations with her mother from age two years and six months to age four years, at least two hours weekly. The child's vowels and some consonants were mostly acquired when data collection started so attention has been paid to consonants that fall within the range defined above. Such consonants include the rhotic (r), the interdentals ( $\theta, \delta$ ), and the velar stops (k, g). It was found that the duration of the rise-age was about four months for all the consonants examined while the start of the rise-age varied from age three years and six months for the interdentals and velar stops to age three years and nine months for the rhotic. What is significant is the duration of the rise-age which should be about the same for typically developing children and not the start of the rise-age which is known to

vary even for typically developing children. It is aimed that that the result of the present study will motivate similar studies for other children so that rise-age norms can be established.

## Probability and Gaussian Stochastic Process applied in Engineering

**Wilker C. Sousa<sup>1</sup>, Natália K. M. Galvão<sup>2</sup>, Fernando F. de Souza<sup>3</sup>,  
Regina C. B. da Fonseca<sup>4</sup>**

*1,2,3,4 Department of Areas IV, Federal Institute of Goiás, Goiânia, Goiás, Brazil, <sup>4</sup>Department of Areas II, Federal Institute of Goiás, Goiânia, Goiás, Brazil,*

*<sup>4</sup>Physics Institute, Darcy Ribeiro Campus, Asa Norte, University of Brasília, Federal District, Brazil*

A stochastic process is a probabilistic mathematical model used for the study of random phenomena that evolve over time. For each moment at time, the random variable has one distribution of probability. To know the whole process normally is used the stochastic methods of Probability Theory. There are many applications of stochastic processes in different areas of knowledge as Physics, Mathematics, Economy and Engineering. To show application in engineering this research has used noises collected from switched circuits, in this case two no-breaks (Exontec UPS 600 and Thor World WEG), because noise plays an important role in problems that involves these types of circuits. Furthermore, there is no mathematical expression capable to define them and they may not be predict at time neither after detected. Presence of noises in signal transference systems is related with the disordered nature of the environment that it advances. To analyze the behavior and type of noises is used the Theory of Probability and the study of stochastic process, analyzing the stochastic moments of noises. The main goal of this research has been describe the importance of the Theory of Probability and stochastic process's in engineering problems, in this case analyzes has been made in the noises series collected in two no-breaks which shows the importance of being detected in the study of the problem, the interferences (internal and external) which influences in the expected results.

**Keywords:** stochastic process, noise, moments of random variable

## Performance measures in discrete supervised classification

Ana Sousa Ferreira<sup>1</sup>, Anabela Marques<sup>2</sup>

<sup>1</sup>*Faculdade de Psicologia, Universidade de Lisboa and Business Research Unit (BRU-IUL) Lisboa, Portugal,* <sup>2</sup>*ESTBarreiro, Setúbal Polytechnic, Portugal and CIIAS-ESS*

The evaluation of results in Cluster Analysis frequently appears in the literature, and a variety of evaluation measures have been proposed. On the contrary, in supervised classification, particularly in the discrete case, the subject of results evaluation is relatively rare in the literature of the area and a part of the measures that have been proposed by some classification researchers are based on many of the measures used in Cluster Analysis. This is the motto for the present study. The evaluation of the performance of any model of supervised classification is, generally, based in the number of cases correctly and incorrectly predicted by the model. However, these measures can lead to a misleading evaluation when data is not balanced. More recently, another type of measures had been studied as coefficients of association or agreement, the Kappa statistics, the Huberty index, Mutual Information or even ROC curves. Exploratory studies have been made to understand the relationship between each measure and data characteristics, namely, samples size, balance and classes' separation. For this purpose, we resort to real and simulated data and use a generalization of the Tobit regression model on the performance of the models.

**Keywords:** Balanced classes, Class separability, Performance measures, Supervised classification

### References:

- Sousa Ferreira, A. and Cardoso, M. G. (2013). Evaluating Discriminant Analysis Results. In *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications* (pp. 155-162). Springer, Berlin, Heidelberg.
- Santafe, G., Inza, I., and Lozano, J. A. (2015). Dealing with the evaluation of supervised classification algorithms. *Artificial Intelligence Review*, 44(4), 467-508.

## Are there Limits for Parameter Settings in Choice-Based Conjoint Analysis?

Winfried J. Steiner<sup>1</sup>, Maren Hein<sup>1</sup>, and Peter Kurz<sup>2</sup>

<sup>1</sup>*Clausthal University of Technology, Department of Marketing, Clausthal-Zellerfeld, Germany,* <sup>2</sup>*bms marketing research + strategy, München, Germany*

Today, choice-based conjoint (CBC) is the most widely used variant of conjoint analysis for collecting and analyzing consumer preferences in marketing research. Its widespread use can be attributed to the development of Hierarchical Bayes (HB) estimation procedures in the mid-1990s, which now allow researchers to account for heterogeneity in consumers' choice behavior at the individual respondent level. In this research, we conduct a simulation study to analyze the capabilities of the HB Logit Model for CBC studies. In particular, we examine how few respondents, how few choice tasks, or how many attributes one can consider in a HB-CBC model before its statistical model performance considerably suffers. Statistical model performance is evaluated under varying factor level settings using criteria for parameter recovery, goodness-of-fit, and predictive accuracy. Our results show that for simple CBC settings HB estimation proves to be quite robust. One of the main findings for simple CBC settings is that holding other factors at convenient levels far more attributes than previously suggested can be used in CBC studies. Further, sample size and/or the number of choice tasks per respondent can be noticeably reduced. However, for more complex CBC settings with an already high number of parameters (part-worths) to be estimated but rather little individual information available from respondents, the HB model is starting to collapse if more than one of those factors (attributes, sample size, choice tasks) is set to an extreme level. Our findings also provide guidance for market researchers who are confronted with the problem that more and more attributes are requested in real-world conjoint settings while the choice task should be kept manageable.

**Keywords:** Choice-Based Conjoint Analysis, Hierarchical Bayes, Simulation

## **Variability and the latent ageing process in life histories: Developing the statistical toolbox**

**David Steinsaltz, Maria Christodoulou**

*Department of Statistics, University of Oxford*

This is a golden age for longitudinal data. Lacking adequate statistical tools to link these vast data collections to core conceptual models, ageing science has struggled to take advantage of these riches. Our work is focused on operationalising core models, the “latent Markov process” models, from the mathematical theory of ageing by means of recently developed statistical methodologies. A key element of many theoretical treatments of ageing is a hidden “senescence” or “vitality” trait that determines an organism’s response to shocks and challenges, its likelihood of reproducing, and its mortality rate. We are applying sequential Monte Carlo (SMC) methods to estimate this process for individuals from complex data. Our main objectives are:

- To parcel out the variability in senescence among time-scales and population scales. In principle, individuals may differ in their initial condition, their inherent rate of ageing, the random shocks from which they suffer, and the age-related deterioration that they accumulate.
- Filtering: We can distil complex longitudinal data into model-based estimates of simple senescence trajectories, that may then be used as the basis for optimal prediction, or as phenotypes for genomic investigations. This talk will outline the general principles of what is possible with these methods, describe the current state of the available statistical tools, and invite consideration and suggestions of types of data and relevant questions that future development of the tools should target.

## **Subset Selection of System Components for Reliability Analysis**

**Eugenia Stoimenova**

*Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia,  
Bulgaria*

We propose applying ranking-and-selection procedures to reliability analysis. That is, we are more interested in whether a given component is better than the others rather than the accuracy of the performance measures. When evaluating  $k$  alternative system designs, one or more

systems are selected as the best; and the probability that the selected systems really are the best is controlled. Let  $m_i$  denote the expected response of system  $i$  and let  $m_{(l)}$  denote the  $l$ th smallest of the  $m_i$  such that  $m_{(1)} \leq m_{(2)} \leq \dots \leq m_{(k)}$ . The goal is to select a subset of size  $m$  containing the  $v$  best of  $k$  systems. We derive the probability lower bound of correctly selecting a subset based on the distribution of order statistics in a clear and concise manner. If  $m = v = 1$ , then the problem is to choose the best system. When  $m > v = 1$ , we are interested in choosing a subset of size  $m$  containing the best. If  $m = v > 1$ , we are interested in choosing the  $m$  best systems. Many selection procedures are derived based on the least favorable configuration (LFC), i.e., assuming  $m_{(1)} = m_{(2)} = \dots = m_{(v)}$  and  $m_{(v)} + d^* = m_{(v+1)} = \dots = m_{(k)}$ . This is because the minimal probability of Correct Selection, occurs under the LFC or the minimal expected losses (risk) from incorrect selection. If the difference  $m_{(v+1)} - m_{(v)} < d^*$ , then these systems are considered to be in the indifference zone for correct selection. On the other hand, if the difference  $m_{(v+1)} - m_{(v)} \geq d^*$ , then these systems are considered to be in the preference zone for correct selection. The goal is to make a correct selection with a probability of at least  $P^*$  provided that  $m_{(v+1)} - m_{(v)} \geq d^*$ . The evaluation of the risk is based on the properties of order statistics.

**Keywords:** Subset selection, Least favorable configuration, Loss functions, Risk estimation

## Statistical estimation in multitype branching processes with multivariate power series offspring distributions

Ana Staneva, Vessela Stoimenova

*Faculty of Mathematics and Informatics, Department of Probability, Operations research and Statistics, Sofia University, Sofia, Bulgaria*

We consider the multitype branching stochastic process with power series offspring distributions. The statistical estimation is carried out under two

sampling schemes - when the entire family tree is observed and when observations only over the generation sizes are sampled. In the case of observable generation sizes we consider a Monte Carlo implementation of the EM algorithm, used as a computationally simple algorithm for numerical approximation of maximum likelihood estimators in incomplete-data problems, which does not require the analytical expression of log-likelihood function. However, the EM algorithm turns out to be slowly convergent in the situations with a significant amount of unobservable data which is the case when one estimates the parameters of multitype branching processes. In order to speed up the convergence we consider an extension of the EM algorithm. The methods are illustrated via simulations and computational results.

**Keywords:** multitype branching processes, multivariate power series distributions, estimation, EM algorithm

### **SIR endemic and epidemic models in random media**

**Mariya Svishchuk<sup>1</sup>, Anatoliy Swishchuk<sup>2</sup>, Yiqun Li<sup>3</sup>**

*<sup>1</sup>Mathematic and Computing Department, Mount Royal University, Calgary, Alberta, Canada, <sup>2</sup>Department of Mathematics and Statistics, University of Calgary, Calgary, Canada, <sup>3</sup>China University of Petroleum, China*

An averaging and diffusion approximation principle for the endemic and epidemic SIR model in a semi-Markov random media is been considered in the following ways:

Numerical examples and their interpretations for two-state Markov and semi-Markov chains.

Two numerical examples involving the data for Dengue Fever Disease (Indonesia and Malaysia (2009)) and Cholera Outbreak in Zimbabwe (2008-2009).

Numerical simulation for the diffusion approximation of the SIR epidemic model in Random Media.



## Assessing labour market mobility in Europe

M. Symeonaki<sup>1</sup>, G. Stamatopoulou<sup>2</sup>

<sup>1,2</sup>*Department of Social Policy, School of Political Sciences, Panteion University of Social and Political Sciences, Athens, Greece*

The present paper proposes a labour market mobility index, that aims at capturing not only the extend of labour market mobility of individuals, but also the quality of their transitions. Different weights for each transition probability are considered given that all transitions are not of equal importance. This is achieved by implementing and testing different weighting scenarios against higher or lower correlation with early job insecurity. More specifically, the proposed index considers only 'positive' transitions and weights transitions between labour market states accordingly, while commonly used mobility indices, take into account either the probability of remaining in the same state or all transition probabilities between states. The proposed methodology is illustrated for the case of young individuals aged between 15 and 29, for the years of the economic crisis, 2008-2016, in European countries using raw data drawn from the EU-LFS survey for those years.

**Keywords:** Mobility index, labour market transitions, labour fluidity, EU-LFS, early job insecurity

## Describing labour market dynamics through Non Homogeneous Markov System theory

M. Symeonaki<sup>1</sup>, G. Stamatopoulou<sup>2</sup>

<sup>1,2</sup>*Department of Social Policy, School of Political Sciences, Panteion University of Social and Political Sciences, Athens, Greece*

The present paper applies non-homogeneous Markov system (NHMS) theory to labour market transitions and provides a cross-national comparison of labour market transitions, among European countries. The paper presents the theoretical adaptation of the NHMS model to labour market dynamics and defines its basic parameters. Raw data drawn from the European Union Labour Force Survey (EU-LFS) is used, in order to estimate and compare the distribution of transition probabilities from the labour market state of employment, unemployment and inactiveness and vice versa, for European countries and examine whether patterns of similar or dissimilar distributions of transition probabilities between labour

market states, exist and for which countries. The paper furthermore reports and compares the school-to-work transition probabilities for European countries.

**Keywords:** labour market transitions, Markov systems, transition probability, EU-LFS

### **Imputation of item non-response in Likert scales using clustering algorithms**

**Maria Symeonaki<sup>1</sup>, Charalampos Papakonstantinou<sup>2</sup>**

*<sup>1</sup>Department of Social Policy, School of Political Sciences, Panteion University of Social and Political Sciences, Athens, Greece, <sup>2</sup>School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece*

In 1932, Likert established a scale for measuring attitudes that has been extensively used in social sciences, educational, medical and health research. Likert scales are comprised of a number of items which are rating scales, and respondents are requested to place themselves on response categories of agreement or disagreement normally scored from 1 to 5 and usually labeled strongly agree, agree, neither agree nor disagree, disagree, strongly disagree. Missing data are often a problem in large-scale surveys, arising when a sampled unit does not respond to the entire survey (unit non-response) or to a particular question (item non-response). It is well accepted that item non-response and the subsequent creation of missing data can harm the quality of measurement, weakening the credibility of the results produced. Many standard statistical techniques and the estimation of the final score, require complete cases and omit subjects with missing data from the study, resulting naturally in loss of information. A solution to this problem is to impute the missing data, i.e. replace the missing values with credible values and consequently create complete data sets, using default solutions provided by the statistical software. However, when dealing with imputation techniques in items forming a Likert scale, one has, among other restrictions, to consider that when simple structure is present, the items altogether measure the given attitude. This means that items are co-related and the imputation technique must take this fact into account. The aim of the present paper is to propose an improved hot-deck type of method for imputing item non-response values in Likert scales and to compare its performance with some well-established procedures. The methodology will be tested to the

CASP-12 scale of measuring quality of life, included in the SHARE project questionnaire.

**Keywords:** Likert scales, attitude measurement, imputation, cluster analysis

### **A neural-network approach for predicting attitudes**

**Maria Symeonaki<sup>1</sup>, Charalampos Papakonstantinou<sup>2</sup>, Catherine Michalopoulou<sup>3</sup>**

*<sup>1,3</sup>Department of Social Policy, School of Political Sciences, Panteion University of Social and Political Sciences, Athens, Greece, <sup>2</sup>National Technical University of Athens, School of Electrical and Computer Engineering, Athens, Greece*

The present paper deals with the application of neural networks techniques to the measurement of attitudes. The methodology provides a way of predicting attitudes based on the available data (respondents' answers to item-questions, questions-indicators and socio-demographic characteristics, such as age, gender and educational level) by dividing the sample randomly in two halves (split-half method). The proposed methodology is illustrated and evaluated on data drawn from a large-scale survey conducted by the National Centre of Social Research of Greece, in order to investigate opinions, attitudes and stereotypes towards the "other" foreigner, and more specifically to the Likert scale developed for the measurement of xenophobia.

**Keywords:** Likert scales, attitude measurement, neural network analysis

### **Aging intensity order**

**Magdalena Szymkowiak<sup>1</sup>**

*<sup>1</sup>Institute of Automation and Robotics, Poznan University of Technology, Poland*

The aging intensity, defined in 2003 by Jiang et al., is a relatively new concept of reliability theory, that can be used in lifetime analysis. Using this function we can determine the aging intensity order, known in the literature as the AI-order (see Nanda et al. 2007). It allows to determine which of the two observed units has a weaker tendency of aging. The support dependent generalized aging intensity of the lifetime random variable has some connections with the support dependent star order

(defined by Danielak and Rychlik in 2003). The relationships between generalized aging intensity order and others stochastic orders is also the subject of our interest (Szymkowiak 2018).

**Keywords:** reliability theory, aging intensity, aging intensity order, star order, stochastic orders

**References:**

- Danielak, K., Rychlik, T. (2003). Sharp bounds for expectations of spacings from DDA and DFRA families. *Statistics and Probability Letters*, 65, 303-316.
- Jiang, R., Ji, P., Xiao, X. (2003). Aging property of unimodal failure rate models. *Reliability Engineering and System Safety*, 79, 113-116.
- Nanda, A.K., Bhattacharjee, S., Alam, S.S. (2007). Properties of aging intensity function. *Statistics and Probability Letters*, 77, 365-373.
- Szymkowiak, M. (2018). Generalized aging intensity functions. *Reliability Engineering and System Safety*, 178, 198-208.

## **A New Method to Relate Multiblock Datasets**

**Essomanda Tchanda-Mangamana<sup>1</sup>, Véronique Cariou, Evelyne Vigneau, Romain Glèlè Kakaï, El Mostafa Qannari**

*<sup>1</sup>PhD Student at StatSC, Oniris, INRA, Nantes, France and LABEF, Université d'Abomey-Calavi, Bénin, Nantes, France*

A new definition of a latent variable associated with a dataset makes it possible to propose variants of the PLS2 regression and the multi-block PLS (MB-PLS). We shall refer to these variants as Rd-PLS regression and Rd-MB-PLS respectively, because they are inspired by both Redundancy analysis and PLS regression. Usually, a latent variable  $t$  associated with a dataset  $Z$  is defined as a linear combination of the variables of  $Z$  with the constraint that the length of the loading weights vector equals 1. Formally,  $t = Z'w$  with  $\|w\| = 1$ . Denoting by  $Z'$  the transpose of  $Z$ , we define herein, a latent variable by  $t = ZZ'q$  with the constraint that the auxiliary variable  $q$  has a norm equal to 1. This new definition of a latent variable entails that, as previously,  $t$  is a linear combination of the variables in  $Z$  and, in addition, the loading vector  $w = Z'q$  is constrained to be a linear combination of the rows of  $Z$ . More importantly,  $t$  could be interpreted as a kind of projection of the auxiliary variable  $q$  onto the space generated by the variables in  $Z$ , since it is collinear to the first PLS1 component of  $q$  onto  $Z$ . Consider the situation in which we aim to predict a dataset  $Y$  from

another dataset  $X$ . These two datasets relate to the same individuals and are assumed to be centered. Let us consider a latent variable  $u = YY'q$  to which we associate the variable  $t = XX'YY'q$ . Rd-PLS consists in seeking  $q$  (and therefore  $u$  and  $t$ ) so that the covariance between  $t$  and  $u$  is maximum. The solution to this problem is straightforward and consists in setting  $q$  to the eigenvector of  $YY'XX'YY'$  associated with the largest eigenvalue. For the determination of higher order components, we deflate  $X$  and  $Y$  with respect to the latent variable  $t$ . Extending Rd-PLS to the context of multi-block data is relatively easy. Starting from a latent variable  $u = YY'q$ , we consider its 'projection' on the space generated by the variables of each block  $X_k$  ( $k=1, \dots, K$ ) namely,  $t_k = X_k X_k' YY'q$ . Thereafter, Rd-MB-PLS seeks  $q$  in order to maximize the average of the covariances of  $u$  with  $t_k$  ( $k=1, \dots, K$ ). The solution to this problem is given by  $q$ , eigenvector of  $YY'XX'YY'$ , where  $X$  is the dataset obtained by horizontally merging datasets  $X_k$  ( $k=1, \dots, K$ ). For the determination of latent variables of order higher than 1, we use a deflation of  $Y$  and  $X_k$  with respect to the variable  $t = XX'YY'q$ . In the same vein, extending Rd-MB-PLS to the path modeling setting is straightforward. Methods are illustrated on the basis of case studies and performance of Rd-PLS and Rd-MB-PLS in terms of prediction is compared to that of PLS2 and MB-PLS.

**Keywords:** Multiblock data analysis, partial least squares regression, path modeling, redundancy analysis

## Robust Estimation for the Single Index Model using Pseudodistances

Aida Toma<sup>1,2</sup>, Cristinca Fulga<sup>3</sup>

<sup>1</sup>Department of Applied Mathematics, The Bucharest University of Economic Studies, Bucharest, Romania, <sup>2</sup>"Gh. Mihoc - C. Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romanian Academy, Bucharest, Romania,

<sup>3</sup>Department of Applied Mathematics, The Bucharest University of Economic Studies, Bucharest, Romania

For portfolios with a large number of assets, the single index model allows to express the large number of covariances between individual asset returns through a significantly smaller number of parameters. This avoids the constraint of having very large samples to estimate the mean and the covariance matrix of the asset returns, which practically would be unrealistic given the dynamic of market conditions. The traditional way to

estimate the regression parameters in the single index model is the maximum likelihood method. Although the maximum likelihood estimators have desirable theoretical properties when the model is exactly satisfied, they may give completely erroneous results when outliers are present in the data set. In this paper we define minimum pseudodistance estimators for the parameters of the single index model and using them we construct new robust optimal portfolios. We prove theoretical properties of the estimators, such as consistency, asymptotic normality, equivariance, robustness, and illustrate the benefits of the new portfolio optimization method for real financial data.

**Keywords:** Minimum Divergence Methods, Robustness, Single Index Model

## **Generalized Lehmann Alternative Type II Family of Distributions and its Application in Record Value Theory**

**Jisha Varghese, K.K. Jose**

*Department of Statistics, St. Thomas College, Mahatma Gandhi University,  
Kottayam, Kerala, India*

In this paper, we introduce and study a new generalized family called Generalized Lehmann Alternative Type II (GLA2) family and introduced special models which include Uniform, Kumaraswamy models under this new family. Generalized Lehmann Alternative Type II Exponential (GLA2E) distribution is also developed and its mathematical properties are obtained along with application. We discuss GLA2E ( $\delta, \beta, \lambda$ ) distributions with special emphasis on record value theory. We derive the entropy of record value distribution and entropy is calculated for various record values.

**Keywords:** Kumaraswamy distribution, Generalized Lehmann Alternative Type II Exponential distribution, Record value, Entropy

### What kind of variables can affect the JIF quartile position of a journal in Dentistry?

**Pilar Valderrama-Baca, Manuel Escabias Machuca, Evaristo Jiménez-Contreras, Mariano J. Valderrama**

*University of Granada, Department of Statistics, Campus Cartuja, Granada, Spain*

This contribution describes an ordinal regression model developed to determine what variables influence the position, by quartiles of the impact factor, of a journal in the field of Dentistry. To this end, 32 journals, 8 pertaining to each quartile were sampled. The estimation procedure concluded that the average number of papers published yearly by a journal and the percentage of systematic reviews are the most significant variables to be considered, along with the factor representing the journal's degree of adherence to recommendations by the International Committee of Medical Journal Editors.

**Keywords:** Dentistry, quartile, journal impact factor, systematic review, ICMJE, ordinal regression

### Application of Markov chain process to predict the natural progression of diabetic retinopathy among adult diabetic retinopathy patients in the coastal area of South India

**Senthilvel Vasudevan<sup>1</sup>, Sumathi Senthilvel<sup>2</sup>, Jayanthi Sureshabu<sup>3</sup>**

*<sup>1</sup>Assistant Professor of Statistics, Department of Pharmacy Practice, College of Pharmacy, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia, <sup>2</sup>Formerly Assistant Professor in Nursing, Amrita College of Nursing, Amrita Vishwa Vidyapeetham, AIMS-Ponekkara, Kochi, Kerala, South India, <sup>3</sup>Formerly Tutor in Medical Entomology, Pondicherry Institute of Medical Sciences, Kalapet, Puducherry, South India*

**Objective:** To find the various stages of DR by using Markov chain analysis Approach and to find the transition of DR. **Materials and Methods:** We have done a retrospective study in the Aravind Eye Hospital, Thavalakuppam, Puducherry. Type 2 diabetes patients were taken in the period of May - June 2012 by using pre designed and pre tested questionnaire. We have concentrated on the Stages of Diabetic Retinopathy with a sample of 200 DR patients. **Study Method and data**

**collection from the patients:** Various stages of DR patients were collected in January 2011 and the stages of the same patients data in the year January 2012. MS Excel 2007 was used for data entry and for analysis SPSS 16.0 version was used. Markov Chain Model approach was used to find out the transition of DR.

**Results:** Out of 200 patients, 126 (63%) were male and 74 (37%) female. The diabetes patients who had type II diabetes for at least five years, a mean age of  $58.80 \pm 10.53$  years and ranged in the age from 27 to 91 years. In one year transition, the probability of an individual in grade-I to move to grade-II is 0.82 which is very high. In the case of the Transition Probability Matrix (TPM) after a period of 5 years it is observed that, the chance of moving from the other lower grades to the final grade is also fairly high.

**Conclusion:** In future, to study the transition of diabetic retinopathy should consider a matrix of estimated transition probabilities, depending on the population, to judge probabilities of transition between states of retinopathy, for the two groups taken up for study and comparison.

**Keywords:** diabetic retinopathy, multi stages, Markov Chain analysis, Puducherry

## Multitype branching processes in random environment

Vladimir Vatutin<sup>1</sup>, Elena Dyakonova<sup>2</sup>, Vitali Wachtel<sup>3</sup>

<sup>1,2</sup>*Department of Discrete Mathematics, Steklov Mathematical Institute, Moscow, Russia,* <sup>3</sup>*Institut für Mathematik, Universität Augsburg, Augsburg, Germany*

Branching processes in random environment (BPRE) serve as models describing the reproduction of particles or individuals within a collective or a population. There are two types of randomness that are incorporated into such branching models: demographical and environmental. Demographical stochasticity means that different individuals give birth independently and their offspring distributions coincide within generations. Environmental stochasticity means that these offspring distributions may change at random from one generation to the next. The basic questions for BPRE are the asymptotic behavior of the survival probability of a population and the rate of growth of the population given its survival. The recent monograph by G. Kersting, V. Vatutin "Discrete Time Branching Processes in Random Environment", ISTE & Wiley, 2017, presents main results obtained up to now by many authors for the single-type BPRE.



Multitype BPPE are less investigated. The point is that their properties are described in terms of products of random matrices whose theory not always contains results suitable to be used for studying multitype BPPE. In this talk we present limit theorems describing the asymptotic behavior of the survival probabilities of the critical and subcritical multitype BPPE and the distribution of the number of particles in a subcritical multitype BPPE given its survival.

Acknowledgement. This work was supported by the Russian Science Foundation under the grant 17-11-01173 and was fulfilled in the Novosibirsk state university.

**Keywords:** multitype branching processes, random environment, survival probability, conditional limit theorem

### **Research of retrial queuing system with called applications in diffusion environment**

**Viacheslav Vavilov**

*Department of Software Engineering, National Research Tomsk State University, Tomsk, Russian Federation*

In this paper, we consider a retrial queue system, where incoming fresh calls arrive at the server according to a Poisson process. Upon arrival, an incoming call either occupies the server if it is idle or joins an orbit if the server is busy. From the orbit, an incoming call retries to occupy the server and behaves the same as a fresh incoming call. After some idle time, the server makes an outgoing call to outside. The system operates in a random environment. Random external factors affect the service time of applications. The mathematical model of a random environment is a diffusion process. For that system we obtained probability distribution of the states of the server and probability distribution of a number of calls in the system.

**Keywords:** Retrial queue, queuing system, random environment, diffusion process, incoming and outgoing calls

### **Clustering of variables approach in a supervised context**

**Evelyne Vigneau<sup>1</sup>**

<sup>1</sup>StatSC, ONIRIS, INRA, Site de la Géraudière, Nantes cedex 3, France

In many areas, advances in data collection techniques make it possible to gather large datasets. The processing of such data poses serious problems due to the large number and the high redundancy between the variables. Within a predictive context, the strategies adopted to overcome this difficulty range from variables selection techniques to regularization methods. Variables selection strategies reduce the dimensionality of the problem and improve the predictive capacity of models, but do not explicitly provide information on the correlation between the selected variables and the other exploratory variables. Our objective is to adapt the CLV approach (Clustering of Variables around Latent Variables, Vigneau and Qannari[1]) to a supervised context, in order to build a predictive model, in a forward and groupwise fashion, with the aim of enhancing the interpretability of the model. The suggested algorithm is a boosting-like procedure for which the base-learner model is constructed from the hierarchical clustering of the exploratory variables. Iteratively, a set of CLV latent components is selected at each hierarchical level. The largest cluster of the exploratory variables, which fulfills a unidimensionality criterion, is finally retained. The residuals of the response variable is regressed on the latent component associated with the retained cluster and the predicted response is updated with the shrunk version of this local predictor. The predictive ability, as well as the interest from the point of view of interpretation, of such an approach will be illustrated in the context of an authentication study of fruit juice mixtures characterized by magnetic resonance spectroscopy (Vigneau and Thomas [2]).

**Keywords:** Clustering of variables, prediction, boosting regression

**References:**

1. E. Vigneau, E. and E.M. Qannari. Clustering of variables around latent components. *Comm. Stat. - Simul Comput.* 32, 4, 1131–1150, 2003.
2. E. Vigneau and F. Thomas. Model calibration and feature selection for orange juice authentication by <sup>1</sup>H NMR spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 117, 22–30, 2012.

## Probabilistic Preference Learning via the Mallows rank model: advances and case studies

Valeria Vitelli<sup>1</sup>, Øystein Sørensen<sup>2</sup>, Elja Arjas<sup>3</sup> and Arnoldo Frigessi<sup>4</sup>

*1 Oslo Center for Biostatistics and Epidemiology, Department of Biostatistics, University of Oslo, Sognsvannveien 9, 0317 Oslo, Norway*

*2 Center for Lifespan Changes in Brain and Cognition, Department of Psychology, University of Oslo, Forskningsveien 3A, 0373 Oslo, Norway*

*3 Department of Mathematics and Statistics, University of Helsinki, Yliopistonkatu 4, 00014 Helsinki, Finland*

*4 Oslo Center for Biostatistics and Epidemiology, Department of Biostatistics, University of Oslo, and Research Support Center, Oslo University Hospital, Sogn Arena, Klaus Torgårds vei 3, 3. etg, 0372 Oslo, Norway*

Ranking items is crucial for collecting information about preferences in many areas, from marketing to politics. The interest often lies both in producing estimates of the consensus ranking of the items, which is shared among users, and in learning individualized preferences of the users, useful for providing personalized recommendations. In the latter task, it is particularly relevant to have posterior distributions of individual rankings, since these can provide an evaluation of the uncertainty associated to the estimates, and thus they can avoid unnecessarily spamming the users.

I will present a statistical model which works well in these situations, and which is able of flexibly handling quite different kind of data. The Bayesian paradigm allows a fully probabilistic analysis, and it easily handles missing data and cluster estimation via augmentation procedures.

Interestingly, this Bayesian framework has also proved to be useful for genomic data integration, since typically heterogeneous microarray data are available from different sources, and their combination allows both to gain statistical power and to strengthen the biological insight.

**Keywords:** Mallows model, Bayesian computing, recommender systems, data augmentation.

## **An exact polynomial in time solution of the one-dimensional bin-packing problem**

**Vassilly Voinov<sup>1</sup>, Rashid Makarov<sup>2</sup>, and Yevgeniy Voinov<sup>3</sup>**

<sup>1,2</sup>*KIMEP University, Almaty, Republic of Kazakhstan, <sup>3</sup>Kaspi Bank, Almaty, Republic of Kazakhstan*

An exact (explicit) polynomial in time algorithm that enumerates all existing optimal solutions of a one-dimensional bin-packing problem is presented. On the contrary of numerous well known heuristic approaches the algorithm is based on the theoretical (polynomial in time) enumeration of the all non-negative integer solutions of a linear Diophantine inequality of any dimension suggested by Voinov and Nikulin in 1997. All combinatorial discrete optimization problems are reduced to obtaining non negative integer solutions of a linear Diophantine inequality and, hence, can be solved in polynomial time. The results can be considered as an empirical proof of the equality  $NP=P$ .

**Keywords:** Combinatorial Optimization, Bin-packing Problem, Linear Diophantine Equation, Column Generation, Polynomial in Time Algorithm..

### **References:**

Voinov V., Nikulin M. (1997) On a subset sum algorithm and its probabilistic and other applications. Balakrishnan N., ed. *Advances in Combinatorial Methods and Applications to Probability and Statistics*. Boston: Birkhäuser, pp. 153-163.

## **Fuzzy data analysis and fuzzy directional data**

**Norio Watanabe**

*Chuo University, Bunkyo-ku, Tokyo, Japan*

A data set is called fuzzy data when each data is represented by a fuzzy set or membership function. Statistical analysis of fuzzy data has been considered by combining statistical methods and the fuzzy set theory. For example, the average fuzzy set can be calculated by applying the extension principle in a natural way. The variance or correlation coefficient can be obtained in the same way. However, the extension principle would lead insignificant results sometimes. Thus the statistical treatment of fuzzy data should be discussed well from a practical viewpoint. In this study we consider the basic statistics for fuzzy data first. Secondly, we focus on

fuzzy directional data. An example of fuzzy directional data is data related to the color circle. The definition of the average fuzzy set for usual fuzzy data is not appropriate for fuzzy directional data. A new definition of the average fuzzy set for fuzzy directional data is introduced. An application is demonstrated by using real data.

**Keywords:** fuzzy data directional data extension principle

### **Performance Comparison of Penalized Regression Method in Logistic Regression under High-dimensional Sparse Data with Multicollinearity**

**Warangkhan Watcharasatian<sup>1</sup>, Supranee Lisawadi<sup>2</sup>, Benjamas Tulyanitikul<sup>3</sup>**

*<sup>1,2,3</sup>Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani, Thailand*

In this paper, penalized regression estimators are proposed to estimate the parameter in logistic regression model with high-dimensional sparse data and high-correlation. Three estimators are considered: Ridge regression, LASSO, and Adaptive LASSO. These estimators that have the ability to solve the multicollinearity problem when the data have high dimension. They are used to compare the performance in term of mean of prediction mean square error (mPMSE) by Monte Carlo simulation on a hundred replicated. The result showed that the Adaptive LASSO estimator has the lowest mPMSE. All in all, Adaptive LASSO performed better than Ridge regression and LASSO.

**Keywords:** Penalized regression, High-dimensional data, High-correlation, Ridge regression, LASSO, Adaptive LASSO

### **FCFS Dynamic Matching Models**

**Gideon Weiss**

*The University of Haifa*

Parallel service systems have several types of customers (participants, recipients), and several types of servers (agents, donors) that arrive along time and are matched according to a compatibility graph, with a focus on First come first served (FCFS) matching. Applications to call centers,

organ transplants, auctions and markets, occur widely. As a queuing system this is generally a highly intractable model, but under Poisson exponential assumption it has partial balance and product form behavior. This leads to the simplified abstraction of FCFS matching of i.i.d. types subject to compatibility graph, where a very complete theory can be obtained. I will discuss this and some open problems that arise with many server scaling.

### **Modeling Sporadic Event Dynamics with Markov-Modulated Hawkes Processes**

**Jing Wu, Tian Zheng**

*Columbia University, New York City, United States*

Modeling event dynamics is central to many disciplines. In particular, point processes models have been applied to explain patterns seen in event arrival times. Such data often exhibits heterogeneous and sporadic trends, which is challenging to conventional methods. It is reasonable to assume that there exists a hidden state process that drives different event dynamics at different states. In this paper, we propose a Markov Modulated Hawkes Process (MMHP) model and develop corresponding inference algorithms. Numerical experiments using synthetic data and data from an animal behavior study demonstrate that MMHP with the proposed estimation algorithms consistently recover the true hidden state process in simulations, and separately captures distinct event dynamics with interesting social structure in real data.

**Keywords:** Bayesian inference, Event dynamics, Hawkes processes, Latent Markov processes

### **Firm Technology Adoption: Optimal Timing and Employee Incentives**

**Yuqian Xu**

*University of Illinois*

With the recent technology boom (i.e., AI, blockchain, etc.), firm managers are facing the strategic issue on how to successfully implement the

innovative technology within the firm. In this paper, we study a firm's strategic decision on innovative technology adoption in a principal-agent setting, with the focus on optimal timing and employee incentives. We consider two scenarios: one with employee incentive misalignment, and the other with alignment. Employees are typically paid through piece rates, and hence if the new technology decreases the piece rate, then they have no incentive to adopt such technology. In this scenario, we characterize the firm's decision on incentive wage contract to motivate its employees to implement the technology. When the new technology increases the piece rate, then the incentives of employees and the firm are aligned with each other, and we characterize the optimal timing of the firm to adopt such technology.

## **Under-five mortality in India: An application of multilevel cox proportional hazard model**

**Awdhesh Yadav**

*International Institute for Population Sciences, Department of Mathematical Demography & Statistics, Mumbai, India*

Despite decline in child mortality in India, under five mortality (U5MR) remains high in India. About 128 million of India's 1.2 billion populations are aged less than 5 years. Although, U5MR of India showed an impressive decline by 9%, a 4 points decline from 43 per 1000 in 2015 to 39 in 2016. The rate of decline has doubled over the last year. In India, more than half of the child deaths occur in the first month of life, with the major clinical causes being complications of prematurity and of delivery. Infectious diseases remain important causes of death both in the first month of life and up to five years of age. Also, Disparities in child health between and within countries have persisted and widened considerably during the last few decades (Bryce, et al., 2006). It is well recognized that disparities in child health outcomes may arise not only from differences in the characteristics of the families that children are born into but also from differences in the socioeconomic attributes of the communities where they live (Fotso & Kuate-Defo, 2005) (Kravdal, 2004) (Ladusingh & Singh, 2006) (Sastry, 1996). While researchers have devoted considerable attention to the impact of individual-level factors on child mortality, less is known about how community characteristics affect health outcomes for children, even though they have a prominent role in theoretical models

most notably Mosley and Chen framework (Mosley & Chen, 1984). The present paper takes advantage of the most recent national survey data to reexamine the issue of contextual effects on childhood mortality in India. In doing so, it contributes to the literature that explores the implications of contextual factors for child mortality by examining the effects of community context on the risk of dying before age five, net of the effect of individual factors.

**References:**

- Bryce, J., Terreri, N., Victora, C. G., Mason, E., Daelmans, B., Bhutta, Z. A., Wardlaw, T. (2006). Countdown to 2015: Tracking Intervention Coverage. *The Lancet*, 368(9541), 1067-1076.
- Fotso, J. C., & Kuate-Defo, B. (2005). Measuring socioeconomic status in health research in developing countries: Should we be focusing on households, communities or Both? *Social Indicators Research*, 72, 189-237.
- IIPS and Macro International. (2015-16). National Family Health Survey 4. Mumbai: International Institute for Population Sciences. Retrieved from <http://rchiips.org/NFHS/nfhs4.shtml>
- Kravdal, O. (2004). Child mortality in India: the community-level effect of education. *Population Studies*, 58, 177-92.
- Ladusingh, L., & Singh, H. C. (2006). Place, community education, gender and child mortality in North-East India. *Population Space and Place*, 12, 65-76.
- Lawn, J. E., Costello, A., Mwansambo, C., & Osrin, D. (2007). Countdown to 2015: Will the Millennium Development Goal for Child Survival be Met? *Arch Dis Child*, 92(6), 551-556.
- Mosley, W. H., & Chen, L. C. (1984). An Analytical Framework for the Study of Child Survival in developing countries. *Population and Development Review*, 10, 25-45.
- Sastry, N. (1996). Community characteristics, individual and household attributes, and child survival in Brazil. *Demography*, 33, 211-229.



## The method of moments and its applications in the theory of stochastic processes

**Elena Yarovaya**

*Lomonosov Moscow State University, Department of Probability Theory, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Moscow, Russia*

To prove the limit theorems of the probability theory and the theory of stochastic processes by the method of moments, the key role is played by the conditions that allow, using the moments of a random variable, to assert that the distribution of a random variable is uniquely determined by its moments. If the answer to this question is affirmative, then the random variable is called M-det. One of M-det conditions used in applications was proposed by J. Stoyanov with co-authors. The question about connection between some sufficient conditions for the unique solvability of the problem for positive moments remains open. We compare two known sufficient conditions that widely used in applications and study the difference between them. Moreover, we managed to obtain a new condition of such type. Then we apply the more general of these conditions, the so-called Carleman condition, in proving the limit theorems for a supercritical branching random walk on multidimensional lattices with a few generation centers of particles under different assumptions on the underlying random walk. The underlying random walk may be symmetric or nonsymmetric, with or without a finite variance of a random walk jumps. We assume that the initial number of particles on the lattice is finite. For the limit theorems of such type we prove that the discrete positive spectrum of the evolutionary operator of the mean number of particles is not empty and the leading eigenvalue is simple. For a supercritical branching random walk in non-homogeneous environments we obtain an exponential growth of particle population over the lattice and at every point of the lattice.

The study was supported by the Russian Foundation for Basic Research, project No. 17-01-468.

**Keywords:** Asymptotic Analysis of Complex Stochastic Evolutionary Systems

## **A Markov Model for Product Line Design**

**Wee Meng Yeo**

*University of Glasgow, Glasgow, United Kingdom*

Even though it is time consuming, a recent survey on sixteen product categories by Market Track shows that 80% of the consumers would always compare prices online. With comparison services available on various platforms such as mobile devices to inform real-time purchasing decisions, sellers are turning to creative ways such as tie-in sales to stay profitable and managing inventory obsolescence. We present a new model to mimic a firm's choice for designing product line design strategy when it has reduced pricing power.

**Keywords:** Markov model, pricing, inventory

## **The implications of applying alternative-supplementary measures of the unemployment rate to regions: Evidence from the European Union Labour Force Survey for Southern Europe, 2008-2015**

**Aggeliki Yfanti<sup>1</sup>, Catherine Michalopoulou<sup>1</sup>, Stelios Zachariou<sup>2</sup>**

*<sup>1</sup>Department of Social Policy, Panteion University of Social and Political Sciences, Greece, <sup>2</sup>Stelios Zachariou, European Commission, DG Eurostat - Unit F3, L- Luxembourg*

The unemployment rate is an important indicator with both social and economic dimensions considered to signify a country's social and economic wellbeing. For its measurement the European Union Labour Force Survey (EU-LFS) is using a synthesized economic construct computed according to the International Labour Organization (ILO) conventional definitions of the employed, unemployed and inactive. However, in the literature, the need for using more than one measure especially in recessionary times is emphasized. In this paper, we investigate the implications of applying two broader alternative definitions of the unemployment rate to regions of interest for social policy purposes. The analysis is based on the 2008-2015 datasets of the EU-LFS for Southern Europe: Greece, Italy, Portugal and Spain. Two alternative measures of the unemployment rate are formulated as variations of the ILO conventional definitions. Applying these two measures to the EU-LFS data, the findings show an increase of the official unemployment rate. Also, they reveal an altered distribution of regional disparities. The results

are reported for the age group 15-74 so as to allow for comparability with the ILO conventional definition of unemployment. Although, the changes in the definitions presented do not exhaust all possibilities, the results indicate the need, especially in recessionary times, for implementing alternative measures of the unemployment rate to the EU-LFS in the tradition of the Current Population Survey.

**Keywords:** EU-LFS, alternative-supplementary measures of the unemployment rate, regional distribution

## **Improving the Understanding of Aliasing Interactions for Designed Experimentation using Regression Tree Methods**

**Timothy M. Young, Robert A. Breyer, Terry Liles, Alexander Petutschnigg**

*University of Tennessee, Center for Renewable Carbon Knoxville, TN, USA*

Quantifying the influence of interactions effects in planned experimentation is very important for scientific discovery. One of the challenges in planned experimentation is preselecting the factors and the levels of factors to be investigated. Regression trees (RT) identify hierarchies of interaction effects. RTs are decision trees where pre-defined statistical functions are used to partition the data space into different class regions. RTs when applied to nonhomogeneous data spaces may result in a smaller generalized error for Y (dependent variable) relative to other supervised learning methods. RTs have noteworthy explanatory value in that the visual 'tree structure' of interrelated predictors is a visual tree of interaction effects. A challenge when planning designed experimentation with fractional factorials is deciding on the appropriate aliasing of interaction effects in the model. Another challenge in the planning stage of designed experimentation is determining the levels of factors that will provide useful inference. RTs were used to quantify the interaction effects and determine the levels of factors. The data set had 3,407 records and 198 regressors. The significant split-points with interaction effects were: 'weight set point,' 'core moisture,' and 'pressing time.' These factors with main and two-level interaction effects were used to create a Box-Behnken response surface model (RSM) with two replicates and three center points for a total of 15 experimental runs. The response surface model had nine terms and simulations revealed a maximization of tensile strength for the interaction 'weight set point and 'core moisture' when 'weight set points' ranged

between 2.9 and 3.3 in the presence of a 'core fiber moisture' 9.2% and 9.8%. Optimal results may not have been realized as quickly using designed experimentation without first conducting a RT.

**Keywords:** Regression trees, interaction effect, designed experimentation, fractional factorials, aliasing, maximization, tensile strength, wood composites

## Evaluating gender differences in Greece, 1985-2017

**Konstantinos N. Zafeiris**

*Laboratory of Physical Anthropology, Department of History and Ethnology,  
Democritus University of Thrace, Komotini, Greece*

Mortality transition has moved well ahead in Greece during the last 30 years. During this course, the longevity of its population was in generally improved; however, after the emergence of the economic crisis in 2008 this course was disturbed. It is also known that females live longer than males and that the causes of death differ significantly among the two genders. Thus, the scope of this paper is to analyze the age-specific and cause-specific contributions to the changing gender differences in life expectancy. For that the Arriaga's method was used. Results are indicative of the differential effects of each cause of death on gender and age and reveal the need for the employment of new policies or intensification of the existing ones in order to improve public health.

**Keywords:** Greece, gender gap, Arriaga's method, causes of death

## Mortality developments in Greece from the cohort perspective

**Konstantinos N. Zafeiris<sup>1</sup>, Anastasia Kostaki<sup>2</sup>, Byron Kotzamanis<sup>3</sup>**

*<sup>1</sup>Laboratory of Physical Anthropology, Department of History and Ethnology,  
Democritus University of Thrace. P. Tsaldari 1, 69132-Komotini, Greece*

*<sup>2</sup>Laboratory of Stochastic Modelling and Applications, Department of Statistics,  
School of Information Studies and Technology, Athens University of Economics  
and Business.*

*<sup>3</sup>Laboratory of Demographic and social analyses (Lads), Department of Planning  
and Regional Development, School of Engineering, University of Thessaly.*

Mortality developments in Greece have been detailly analyzed and discussed with the aid of period data. However, any relative studies from

the cohort perspective are absent. Such studies are of great importance because period analysis and analogously the estimation of the relevant life expectancies can give a distorted picture of the real temporal trends of longevity when mortality changes. Tempo and cohort effects as well as other agents like selection are responsible for this phenomenon.

Taking these into consideration, period life table data were used separately for the male and female population in order to obtain one-year probabilities of death for any birth cohort formed in Greece after the 1950s. Afterwards, partial life expectancies and the expected years lost between birth and several other ages were calculated for each of them. The results of the analysis are indicative of the mortality transition observed in Greece in the last 57 years and give a clear picture of the existing gender differences during this transition.

**Keywords:** mortality, cohort analysis, partial life expectancy, expected years lost

### **TURCOSA: User-friendly, Personalized and Cloud-Based Statistical Analysis System**

**Gökmen Zararsız<sup>1</sup>, Selçuk Korkmaz<sup>2</sup>, Dinçer Göksülük<sup>3</sup>, Gözde Ertürk Zararsız<sup>1</sup>, Merve Başol Göksülük<sup>3</sup>, Ahmet Öztürk<sup>1</sup>**

<sup>1</sup>Erciyes University, Faculty of Medicine, Department of Biostatistics, Kayseri, Turkey, <sup>2</sup>Trakya University, Faculty of Medicine, Department of Biostatistics, Edirne, Turkey, <sup>3</sup>Hacettepe University, Faculty of Medicine, Department of Biostatistics, Ankara

In various fields, users collect data to seek solutions for various research problems. These fields include banking, public services, insurance, healthcare, biotechnology, communications, manufacturing, energy, capital markets, and the collected data is analyzed using statistical software. However, the users are experiencing various problems and difficulties while using these statistical software. Firstly, there are many users who do not have enough knowledge of statistical terminology. In addition, some statistical software require coding skills and abilities to carry out certain analysis. Due to these problems, the users are experiencing with problems both in selection and application of the appropriate statistical methods, also the interpretation and reporting of the statistical results. In this study, we developed TURCOSA, a cloud-based statistical analysis software, to simplify the analysis procedures for researchers. The users can access the software via any internet

connected computer regardless of the type, performance and operating system of their devices. Using TURCOSA, users can create projects, upload multiple datasets from different formats and invite their colleagues to analyze their data together using the user-friendly analysis modules of the software. The software includes easy-use modules, which automatically selects and applies the appropriate statistical methods based on the type of the data and the hypothesis of the researchers. Moreover, TURCOSA interprets the analysis results and reports the outputs with interactive tables and graphs. The software can be accessed from the following website, [www.turcosa.com.tr](http://www.turcosa.com.tr). This work was supported by TUBITAK [116E211].

**Keywords:** Biostatistics, cloud informatics, data analysis, data mining, statistical software

## The Rate of Growth and Fluctuations of Compound Renewal Processes

Nadiia Zinchenko<sup>1</sup>

<sup>1</sup>*Department of Information Technology and Data Analysis, Nizhyn State Mukola Gogol University, Nizhyn*

We consider the compound renewal processes of the form  $D(t) =$

$S(N(t)) = \sum_{i=1}^{N(t)} x_i$ , where  $N(t)$  is a renewal process and  $\{X_i\}$  are r.v. independent of  $N(t)$ . Our main task is finding the conditions on summands  $\{X_i\}$  and inter-renewal intervals  $\{Z_i\}$  and the form of normalizing and centering functions  $f(t)$  and  $m(t)$ , for which a.s.  $\limsup_{t \rightarrow \infty} (D(t) - m(t))/f(t) = c_1$  or  $\liminf_{t \rightarrow \infty} (D(t) - m(t))/f(t) = c_2$ ,  $c_1, c_2 = \text{const}$ . Similar problem concerning the rate of growth of increments  $\Delta(t) = \Delta(t, a(t)) = D(t+a(t)) - D(t)$  on intervals, whose length  $a(t)$  grows but not faster than  $t$ , is also discussed. A number of integral tests for investigation of the upper/lower functions for  $D(t)$  and  $\Delta(t)$  under various assumptions on renewal process, moment and dependent conditions of random summands  $\{X_i\}$  are proposed. The cases of independent, weakly dependent and associated summands with finite variance are studied as well as martingales and random variables satisfying  $\phi$ -mixing conditions. Also the case of i.i.d. summands attracted to  $\alpha$ -stable law ( $1 < \alpha < 2$ ) is studied in details. As a consequence various modifications of the LIL and Erdős-Rényi-Csörgő-

Révész-type SLLN for compound renewal processes are obtained and used for investigation fluctuations of the risk processes in classical Cramer-Lundberg and in renewal Sparre Andersen risk models. In this contexts we mainly focused on the problem of large claims. The case of risk processes with stochastic premiums, where both total claim amounts and total premium amounts are compound Poisson (and more general – compound renewal) processes, is also investigated in the same manner. A nice background for our investigation is a number of general results about strong approximation of the compound renewal processes by a Wiener or  $\alpha$ -stable Lévy processes.

**Keywords:** Compound Renewal Process, Strong Limit Theorems, Strong Approximation, Integral Test, Risk Process, Law of Iterated Logarithm

## Groundwater level forecasting for water resource management

**Andrea Zirulia<sup>1,2,3</sup>, Enrico Guastaldi<sup>1,3,4</sup>, Alessio Barbagli<sup>1</sup>**

<sup>1</sup>CGT Center for GeoTechnologies, University of Siena, Via Vetri Vecchi 34, 52027 San Giovanni Valdarno, Italy, <sup>2</sup>Department of Chemical and Geological Sciences, University of Cagliari, Via Trentino 51, Cagliari, Italy, <sup>3</sup>GeoExplorer Impresa Sociale Srl, Arezzo, Italy, <sup>4</sup>CGT Spinoff Srl, Arezzo, Italy

The actual increasing of groundwater demand, which depends on several natural (eg. climate change) and human factors (eg. agriculture, domestic and industrial use), is cause of depletion in both quantity and quality of the groundwater resource. Analysis of the groundwater level time series data and predicting their future trends could be an alternative way, respect to local numerical groundwater models, to manage the water use in large areas, aimed to a sustainable development and useful to identify causes of water level decline. In order to estimate where the groundwater system directs and to show the usefulness of such methodology for decisions of public interest, we have studied a small area located on the Tuscan coast (Italy). Results of the integrated time series analysis not only give informations about future hydrologic trends, but can be also useful to understand possible climate change and related effects in hydrologic system.

**Keywords:** Time series, Groundwater, Water level, Climate change

---

## Table of Contents

---

### Reliability Modelling and Assessment of a Heterogeneously Repaired System with Partially Relevant Recurrence Data

Narayanaswamy Balakrishnan ..... 1

### Approximations with Error Bounds in Applied Probability Models: Exponential and Geometric Approximations

Mark Brown..... 1

### New Filters for the Calibration of Regime Switching BETA Dynamics

Robert J. Elliott, Carlton Osakwe..... 2

### Entropy Rates of Markov Chains

Valerie Girardin ..... 3

### Reinforcement Learning: Connections between MDPS and MAB problems

Michael N. Katehakis ..... 3

### Schur-Constant and Related Dependency Models

Claude Lefèvre..... 4

### Applied Stochastic Models: Theory vs Applications?

Christos H Skiadas..... 4

### Applied and conceptual meaning of multivariate failure rates and load-sharing models

Fabio L. Spizzichino ..... 5

### On the Theory and Applications of Nonhomogeneous Markov Set Systems

P.-C.G. Vassiliou..... 6

### A note on serious arguments in favor of equality $P=NP$

Vassily Voinov..... 6

### Sparse Correspondence Analysis

Ruiping Liu, Ndeye Niang, Gilbert Saporta, Huiwen Wang..... 7



<b>From Process Modelling to Process Mining: using Big Data for Improvement</b>	
Sally McClean .....	7
<b>Asymptotic Algorithms of Phase Space Reduction and Ergodic Theorems for Perturbed Semi-Markov Processes</b>	
Dmitrii Silvestrov .....	8
<b>Structural Equation Modeling: Infant Mortality Rate in Egypt Application</b>	
Fatma Abdelkhalek, Marianna Bolla.....	9
<b>A Topological Multiple Correspondence Analysis</b>	
Rafik Abdesselam .....	9
<b>On PageRank Update in Evolving Tree Graphs</b>	
Benard Abola, Pitso Seleka Biganda, Christopher Engström, John Mango Magero, Godwin Kakuba, Sergei Silvestrov .....	11
<b>Comparison of stability conditions for queueing systems with simultaneous service</b>	
Larisa Afanaseva, Svetlana Grishunina .....	11
<b>Commute times and the effective resistances of random trees</b>	
Fahimnah Alawadhi .....	12
<b>Technical Efficiency of Public Healthcare Systems in Uttar Pradesh and Maharashtra: A Data Envelopment Analysis</b>	
Cheryl Anandas.....	13
<b>Classification Methods for Healthcare System Costs in EU</b>	
A. Anastasiou, P. Hatzopoulos, A. Karagrigoriou, G. Mavridoglou .....	14
<b>On demographic approach of the BGGM distribution parameters</b>	
Panagiotis Andreopoulos Alexandra Tragaki, Fragkiskos G. Bersimis, Maria Moutti .....	14
<b>Shortest parts in Markov-modulated networks</b>	
A.M Andronov, I.M. Dalinger, I.V. Yackiv.....	15

**A comparison of graph centrality measures based on random walks and their computation.**

Collins Anguzu, Christopher Engström, Sergei Silvestrov ..... 15

**Data-driven predictive modelling of enrolment associated processes to optimize clinical trial's operations**

Vladimir Anisimov ..... 16

**Robustness to outlying variables in PCA**

Arlette Antoni, Thierry Dhorne ..... 17

**Floods: Statistics, Analysis, Regulation**

Valery Antonov, Roman Davydov ..... 18

**Fractal analysis of nanostructured material objects**

Valery Antonov, Anatoly Kovalenko ..... 18

**Kernel SVM Distance Based Control Chart for Statistical Process Monitoring**

Anastasios Apsemidis, Stelios Psarakis ..... 19

**Modeling private preparedness behavior against flood hazards**

Pedro Araujo, Gilvan Guedes, Rosangela Loschi ..... 20

**Inflation Rates Indicators and their Properties**

Josef Arlt, Markéta Arltová ..... 21

**A new simplex distribution allowing for positive Covariances**

Roberto Ascari, Sonia Migliorati, Andrea Ongaro ..... 22

**Application of meta-heuristic optimization approaches in operational planning for clinical trials**

Matthew Austin, Stephen Gormley, Vladimir Anisimov ..... 23

**Increasing efficiency in the EBT algorithm**

Tin Nwe Aye,, Linus Carlsson ..... 24

**Clustering of multiple lifestyle risk factors and health-related quality of life in Korean population**

Younghwa Baek, Kyungsik Jung, Hoseok Kim ..... 24

**Multiple outliers identification in linear regression**

Vilijandas Bagdonavičius, Linas Petkevičius..... 25

**Residual based goodness-of-fit tests for linear regression models**

Vilijandas Bagdonavičius, Rūta Levulienė ..... 26

**Structure of the particle field for a branching random walk with a critical branching process at every point**

Daria Balashova, Stanislav Molchanov, Elena Yarovaya..... 27

**Detecting long term and abrupt changes in hydrological processes**

Dominika Ballová, Adam Šeliga ..... 27

**K-optimal designs for parameters of shifted Ornstein–Uhlenbeck processes**

Sándor Baran ..... 29

**Data mining application issues in the taxpayers selection process**

Mauro Barone, Andrea Spingola, and Stefano Pisani ..... 29

**Greed and Fear: the Nature of Sentiment\***

Giovanni Barone-Adesi, Matteo M. Pisati, Carlo Sala ..... 30

**Revised survival analysis-based models in medical device innovation field**

Andrea Bastianin, Emanuela Raffinetti..... 31

**Pre-emergence thermal and hydrothermal time model in crop**

Behnam Behtari..... 32

**Estimating the width of uniform distribution under measurement errors**

Mirta Benšić, Safet Hamedović, Kristian Sabo ..... 33

**Nonparametric Regression Estimator for LTRC and Dependent Data**

Siham Bey, Zohra Guessoum, Abdelkader Tatachak..... 34

**Unimodality and Logconcavity of Density Functions of System Lifetimes**

Mariusz Bieniek, Marco Burkschat, Tomasz Rychlik ..... 35

**PageRank and Perturbed Markov chains**

<b>194</b>	<i>The 18th ASMDA International Conference (ASMDA 2019)</i>	
	Pitos Seleka Biganda, Benard Abola, Christopher Engström, John Mango Magero, Godwin Kakuba, Sergei Silvestrov .....	35
	<b>A Decomposition of Change in Disabled Life Expectancy at Retirement</b>	
	Heather Booth and Qi Cui .....	36
	<b>Efficiency of Brazilian Hospitals: assessment of Unified Health System (SUS)</b>	
	Laura de Almeida Botega, Mônica Viegas Andrade, Gilvan <i>Ramalho Guedes</i> .....	37
	<b>Current multibloc methods. A comparative study in a unified framework</b>	
	Stéphanie Bougeard, Ndèye Niang, Thomas Verron, Xavier Bry .....	39
	<b>A New Modified Scheme for Linear Shallow-Water Equations</b>	
	Aicha Boussaha.....	40
	<b>Redistricting Using Counties, Municipalities and the Convexity Ratio</b>	
	James R. Bozeman .....	41
	<b>Modelling monthly birth and deaths using Seasonal Forecasting Methods as an input for population estimates</b>	
	Jorge Miguel Bravo, Edviges Coelho.....	41
	<b>New Dividends Strategies</b>	
	Ekaterina Bulinskaya.....	42
	<b>Goodness-of-Fit Testing for Point Processes in Survival Analysis</b>	
	Sami Umut Can, Estate V. Khmaladze, Roger J.A. Laeven .....	43
	<b>One bank problem in the federal funds market</b>	
	Elena Cristina Canepa, Traian A. Pirvu.....	44
	<b>Calibration of two-factor stochastic volatility model</b>	
	Betuel Canhanga, Ying Ni, Calisto Guambe .....	45
	<b>The wide variety of regression models for lifetime data</b>	
	Chrys Caroni .....	46
	<b>Quantization of Transformed Lévy Measures</b>	
	Mark Anthony Caruana.....	47

**Modeling of mortality in elderly by trachea, bronchus and lung cancer diseases in the Northeast of Brazil**

João Batista Carvalho, Neir Antunes Paes ..... 47

**On a family of risk measures based on largest claims**

Antonia Castaño, Gema Pigueras, Miguel Ángel Sordo..... 48

**A general piecewise multi-state survival model for the study on the progression of breast cancer**

Juan Eloy Ruiz-Castro and Mariangela Zenga ..... 49

**Performance estimation of a wind farm with a copula dependence structure**

Laura Casula, Guglielmo D’Amico, and Giovanni Masala and Filippo Petroni and Robert Adam Sobolewski ..... 49

**Bayesian analysis for the system lifetimes under Frank copulas of Weibull component lifetimes**

Ping Shing Chan, Yee Lam Mo ..... 51

**Psychometric validation of constructs defined by ordinal-valued items**

Anastasia Charalampi, Catherine Michalopoulou, Clive Richardson ..... 51

**Calculation of Analogs of Lyapunov Indicators for Different Types of EEG Time Series Using Artificial Neural Networks**

German Chernykh, Ludmila Dmitrieva, Yuri Kuperin ..... 52

**Long Term Care Insurance in Singapore: Assessing the Shield Index**

Ngee Choon Chia ..... 53

**Real Estate Market Analysis with Time-frequency Decomposition**

Siu Kai Choy, Tsz Fung Stanley Zel ..... 54

**Variability and the latent ageing process in life histories: Applications on cohort studies**

M.D. Christodoulou, J.A. Brettschneider, D. Steinsaltz, ..... 55

**Polya-Aeppli Geometric process**

Stefanka Chukova, Leda D. Minkova ..... 56

**Latent Class Analysis in the psychological research on attachment styles and the transition to parenthood**

Franca Crippa, Mariangela Zenga, Rossella Shoshanna Procaccia, Lucia Leonilde Carli..... 56

**Alcohol consumption in selected European countries**

Jana Vrabcová, Kornélia Svačinová, Markéta Pechholdová ..... 57

**A proportional hazard model under bivariate censoring and truncation**

Hongsheng Dai, Chao Huang, Miriam J. Johnson, Marialuisa Restaino ..... 58

**Simultaneous Threshold Interaction Modeling Approach for Paired Comparisons Rankings**

Antonio D'Ambrosio, Alessio Baldassarre and Claudio Conversano..... 59

**A Probabilistic Model of Wind Farm Power Generation via Copulas and Indexed Semi-Markov Models**

Guglielmo D'Amico, Giovanni Masala, and Filippo Petroni and Robert Adam Sobolewski ..... 59

**Rocof of higher order for multistate systems in continuous time**

Guglielmo D'Amico and Filippo Petroni..... 60

**A copula based Markov Reward approach to the credit spread in European Union**

Guglielmo D'Amico, Filippo Petroni, Philippe Regnault, Stefania Scocchera, Lorian Storch ..... 61

**The Generalized Calculation of Pure Premium for Non-Life Insurance by Generalized Non-Homogeneous Markov Reward Processes**

Guglielmo D'Amico, Jacques Janssen, Filippo Petroni, Raimondo Manca and Ernesto Volpe Di Prignano ..... 62

**Dynamic inequality: a Python tool to compute the Theil inequality within a stochastic setting**

Guglielmo D'Amico , Stefania Scocchera, Lorian Storch..... 63

**On the number of observations in random regions determined by records**

Anna Dembińska, Masoumeh Akbari, Jafar Ahmadi ..... 64

**Expected lifetimes of coherent systems with DNID components**

Anna Dembińska, Agnieszka Goroncy ..... 64

**Residual lifetime of k-out-of-n: G systems with a single cold standby unit**

Anna Dembińska, Nikolay I. Nikolov, Eugenia Stoimenova ..... 65

**Co-clustering for time series based on a dynamic mixed approach for clustering variables**

Christian Derquenne ..... 66

**Estimating gross margins for agricultural production in the EU: approaches based on the equivariance of quantile regression**

Dominique Desbois ..... 67

**Pricing of Longevity derivatives and cost of capital**

Pierre Devolder, Fadoua Zeddouk ..... 68

**A Flexible Regression Model for Compositional Data**

Agnese M. Di Brisco, Sonia Migliorati ..... 68

**Reverse Mortgages: Risks and Opportunities**

Emilia Di Lorenzo, Gabriella Piscopo, Marile Sibillo, Roberto Tizzano ..... 69

**Introduction of reserves in self adjusting steering the parameters of a pay-as-you-go pension plan**

Keivan Diakite, Abderrahim Oulidi, Pierre Devolder ..... 70

**The Distribution of the Inverse Cube Root Transformation of Error Component of the Multiplicative Time Series Model**

Dike Awa, Chikezie David, Otuonye Eric ..... 71

**Computation of the optimal policy for a two-compartment single vehicle routing problem with simultaneous pickups and deliveries, stochastic continuous demands and predefined customer order**

Theodosios D. Dimitrakos, Constantinos C. Karamatsoukis, Epaminondas G. Kyriakidis ..... 72

**Entropy Analytics of Euro-Disney Facebook Five Star Rating Dataset**

Yiannis Dimotikalis ..... 73

**Application of Modified Local Holder Exponents Method to Study the Multichannel EEG in States of Meditation and Background**

Ludmila Dmitrieva, Yuri Kuperin, German Chernykh, Daria Kleeva ..... 74

**Stochastic Modeling of Affinity Based Cognitive Networks**

Denise Duarte, Gilvan Guedes, Gilvan Guedes, Rodrigo Ribeiro ..... 75

**Comparative study of two different SW-RPA approaches to calculate the excess entropy of some liquid metals**

N.E. Dubinin..... 76

**Correcting death rates at advanced old age: review of models**

Dalkhat M. Ediev..... 76

**Estimating differences between two models based on different input data and environmental factors**

Christopher Engström..... 77

**Cumulative Density Function from Contaminated Noise**

Ben Jrada M Es-salih, Djaballah Khadidja ..... 78

**Identification of thyroid cancer risk factors incidence in urban and rural areas, Pakistan**

Asif Faiza, Muhammad Noor-ul-Amin ..... 79

**Changes of the Tehran City Floating Population Based on the 2006 and 2011 Census Data**

Reza Sotoudeh Farkosh, Ashraf Mashhadi Heidar ..... 80

**Predicting The Customers Trend in Digital Firms Case Study in Iran**

Saeed Fayyaz ..... 81

**Is Taylor's power law true for random networks?**

István Fazekas, Csaba Noszály, and Noémi Uzonyi ..... 82

**Design of clinical trials with “time-to-event” end points and under Poisson-gamma enrollment model**

Valerii Fedorov ..... 83



**Healthy Ageing in Czechia**

Tomas Fiala, Jitka Langhamrova ..... 83

**Construction and Universal Representation of k -Variate Survival Functions**

Jerzy Filus, Lidia Filus ..... 84

**One step-ahead predictive ability in nested regression models**

S.B. Fotopoulos, S. Lyu, V.K. Jandhyala, A. Kaul ..... 85

**Poisson regression and change-point analysis**

Jim Freeman, Yijin Wen ..... 86

**Two-way cross balanced ORDANOVA**

Tamar Gadrich ..... 86

**Robust Minimal Markov Model**

Jesús E. García, V.A. González-López ..... 87

**Stochastic Profile of Strains of Zika from Tropical and Subtropical Regions**

Jesús E. García, V.A. González-López, S.L. Mercado Londoño, M.T.A. Cordeiro ..... 88

**Generating a ranking on a set of alternatives from the qualitative assessments given by agents with different expertise**

José Luis García-Lapresta, and Raquel González del Pozo ..... 89

**Error Detection in sequential laser sensor input**

Gwenaël Gatto, Olympia Hadjiliadis ..... 90

**Prediction intervals for weighted TAR forecasts**

Francesco Giordano, Marcella Niglio ..... 90

**Filling the Gap between Continuous Time Autoregressive Processes and Discrete Observations**

Valrie Girardin, Rachid Senoussi ..... 91

**Maximization problem subject to constraint of availability in semi-Markov model of operation**

Franciszek Grabski ..... 92

<b>Prediction of the 2019 IHF World Men's Handball Championship -- An underdispersed sparse count data regression model</b>	
Andreas Groll, Jonas Heiner, Gunther Schauburger, Jörn Uhrmeister .....	92
<b>Health vulnerability related to climate extremes in Amazonia and the Brazilian Northeast</b>	
Gilvan Guedes, Pollyane Silva, Maria Helena Spyrides, Claudio Silva, Lara Andrade, Kenya Noronha.....	93
<b>Generational differences in health-related quality of life among Brazilian gay men</b>	
Gilvan Guedes, Samuel Araujo, Paula Ribeiro, Kenya Noronha .....	94
<b>Likelihood-comparison of alternative Markov models incorporating duration of stay</b>	
Marie-Anne Guerry, Philippe Carette.....	95
<b>Fractional Difference ARFIMA Models for long memory timesies</b>	
Maryam Haghiri.....	96
<b>Minimizing Expected Discounted Cost in Queueing Loss Models with Discriminating Arrivals</b>	
Babak Haji .....	96
<b>Estimation of the relative error in regression analysis under random left-truncation model</b>	
Farida Hamrani .....	97
<b>Robust Regression in Time Series under Truncated and Censored data</b>	
Benseradj Hassiba, Guessoum Zohra .....	97
<b>Sequential on-line detection and classification in 3D Computer Vision</b>	
Olympia Hadjiliadis .....	98
<b>Cone distribution functions and quantiles for multivariate random variables</b>	
Andreas Hamel, Daniel Kostner .....	98
<b>Introducing and evaluating a new multiple component stochastic mortality model</b>	

P. Hatzopoulos, A. Sagianou .....	99
<b>Modeling of the extreme and records values for precipitation and temperature in Lebanon</b>	
Ali Hayek*, Nabil Tabaja, Zaher Khraibani, Samir Abbad Andaloussi, Joumana Toufaily, Evelyne Garnie-Zarli, Tayssir Hamieh* .....	100
<b>Births by order and childlessness in the post-socialist countries</b>	
Filip Hon, Jitka Langhamrova .....	101
<b>Brand-Level Market Basket Analysis by Conditional Restricted Boltzmann Machines</b>	
Harald Hruschka .....	101
<b>Death, Disease, Failure Prediction: Survival Models vs Statistical Machine Learning/Reduced Order Models</b>	
Catherine Huber-Carol .....	102
<b>Solving Rank Aggregation Problems through Memetic Algorithms</b>	
Carmela Iorio, Giuseppe Pandolfo, Autilia Vitiello .....	103
<b>Health status and social activity of men and women at pre- and retirement age in Russia</b>	
Alla Ivanova, Elena Zemlyanova, Tamara Sabgaida, Sergey Ryazantsev .....	104
<b>Optimising Group Sequential and Adaptive Clinical Trial Designs: Where Frequentist meets Bayes</b>	
Christopher Jennison .....	105
<b>A Factor Analysis of Factor Shares, Price Rigidities and the Inflation-Output Trade-Off</b>	
Christian Jensen .....	105
<b>American option pricing under a Markovian regime switching model</b>	
Lu Jin, Yuji Sakurai, Ying Ni .....	106
<b>Revisiting Transitions between Superstatistics</b>	
Petr Jizba, Martin Prokš .....	106
<b>Generalized T-X family of distributions and their applications</b>	

K. K. Jose, Jeena Joseph .....	107
<b>Distributionally Robust Optimization with Data Driven Optimal Transport Cost and its Applications in Machine Learning</b>	
Yang Kang, Jose Blanchet, Fan Zhang, Karthyek Murthy.....	108
<b>A Joint Modelling Approach in SAS to Assess Association between Adult and Child HIV infections in Kenya</b>	
Elvis Karanja, Naomi Maina, June Samo .....	108
<b>Real time prediction of infectious disease outbreaks based on Google trend data in Africa</b>	
Elvis Karanja .....	109
<b>Multivariate Random Sums: Limit Theorems, Related Distributions and Their Properties.</b>	
Yury Khokhlov, Victor Korolev.....	110
<b>Investigating some attributes of periodicity in DNA sequences via semi Markov modelling</b>	
Pavlos Kolias and Aleka Papadopoulou .....	111
<b>Spatio-temporal Aspects of Community Well-Being In Multidimensional Functional Data Approach</b>	
Mirosław Krzyśko, Włodzimierz Okrasa, Waldemar Wołyński .....	111
<b>Mixed Fractional Brownian Motion</b>	
Kęstutis Kubilius, Aidas Medžiūnas .....	113
<b>Optimal collection of two materials from N ordered customers with stochastic continuous demands</b>	
Epaminondas G. Kyriakidis, Theodosios D. Dimitrakos, Constantinos C. Karamatsoukis .....	113
<b>Identifying the characteristics influencing the mathematical literacy in Spanish students</b>	
Ana María Lara-Porras, María del Mar Rueda-García, David Molina-Muñoz .....	114
<b>I-Delaporte process and applications</b>	

Meglana D. Lazarova, Leda D. Minkova.....	115
<b>Mobile learning for training bioinformatics in the connected world</b>	
Taerim Lee, Juhan Kim .....	116
<b>Data Fraud and Inlier Detection</b>	
H.-J. Lenz, W. Kössler .....	117
<b>Balancing Covariates in Regression Discontinuity Designs</b>	
Shuangning Li .....	117
<b>Using the Developing Countries Mortality Database (DCMD) to Probabilistically Evaluate the Completeness of Death Registration at Old Ages</b>	
Nan Li, Hong Mi .....	118
<b>Gaussian Limits for Multichannel Networks with Input Flows of General Structure</b>	
Hanna V. Livinska, Eugene O. Lebedev .....	119
<b>A cluster analysis of multiblock datasets</b>	
Fabien Llobell, Véronique Cariou, Evelyne Vigneau, Amaury Labenne, El Mostafa Qannari.....	119
<b>Properties of the Hardlims*Tansig Model of the Statistical Neural Network</b>	
Olamide O. Ilori, Christopher G. Udomboso .....	121
<b>Properties of the extreme points of the probability density distribution of the Wishart matrix</b>	
Karl Lundengård, Asaph Keikara Muhumuza, Sergei Silvestrov, John Mango, Godwin Kakuba .....	122
<b>Comparison of parametric models applied to mortality rate forecasting</b>	
Karl Lundengård, Samya Suleiman, Hisham Sulemana, Milica Rancic, Sergei Silvestrov .....	122
<b>Particle filter impoverishment under an urn model perspective</b>	
Rodi Lykou, George Tsaklidis.....	123
<b>Bayesian model for mortality projection: evidence from Central and Eastern Europe</b>	

<b>204</b>	<i>The 18th ASMDA International Conference (ASMDA 2019)</i>	
	Justyna Majewska, Grazyna Trzpiot .....	123
	<b>Itô Type Bipartite Fuzzy Stochastic Differential Equations with Osgood condition</b>	
	Marek T. Malinowski .....	124
	<b>Asymptotics of Implied Volatility in the Gatheral Double Stochastic Volatility Model</b>	
	Mohammed Albuhayri, Anatoliy Malyarenko, Sergei Silvestrov, Ying Ni, Christopher Engström, Finnan Tewolde, Jiahui Zhang .....	125
	<b>Latent class detection in Latent Growth Curve Models</b>	
	Katerina M. Marcoulides, Laura Trinchera .....	126
	<b>Properties of patterns in a semi-Markov chain</b>	
	Brenda Ivette Garcia Maya and Nikolaos Limnios .....	126
	<b>Generalized First Passage Time Method for the Estimation of the Parameters of the Stochastic Differential Equation of the Black-Scholes Model</b>	
	Samia Meddahi, Khaled Khaldi.....	127
	<b>Skorokhod embeddings in Brownian Motion and applications to Finance</b>	
	Isaac Meilijson .....	128
	<b>Kernel estimator regression in censored and associated models</b>	
	Nassira Menni, Abdelkader Tatachak.....	128
	<b>Discrete Time Risk Model</b>	
	Leda D. Minkova .....	129
	<b>Multichannel sequence analysis to identify patient pathway</b>	
	Elisabeth Morand .....	129
	<b>Determining influential factors in spatio-temporal models</b>	
	Rebecca Nalule Muhumuza, Olha Bodnar, Sergei Silvestrov, Joseph Nzabanita, Rebecca Nsubuga.....	130
	<b>A partial unemployment among youths in India owing to under reporting of age of older cohorts</b>	

Barun Kumar Mukhopadhyay .....	131
<b>A Comparative Study for Forecasting Stochastic Volatility Models: EWMA model versus Heston model</b>	
Jean-Paul Murara, Anatoliy Malyarenko, Milica Rancic, Ying Ni, Sergei Silvestrov .....	132
<b>Forecasting Stochastic Volatility for Exchange Rate</b>	
Jean-Paul Murara, Anatoliy Malyarenko, Milica Rancic, Ying Ni, Sergei Silvestrov .....	132
<b>Pricing Options under two-dimensional Black-Scholes Equations by using C-N Scheme</b>	
Jean-Paul Murara, Anatoliy Malyarenko, Ying Ni, Sergei Silvestrov .....	133
<b>A Mathematical Model for Cannibalism in a Predator-Prey System with Harvesting</b>	
Loy Nankinga .....	134
<b>Stochastic Modeling for Weather Derivatives and Application to Insurance</b>	
Clarinda Nhangumbe, Alex Marrime, Betuel Canhanga, Calisto Guambe ..	134
<b>Goodness-of-fit tests for logistic family via characterization</b>	
Yakov Nikitin, Ilya Ragozin .....	135
<b>An Algebraic Method for Pricing Financial Contracts in the Post-Crisis Financial Market</b>	
Hossein Nohrouzian, Anatoliy Malyarenko, Ying Ni, Christopher Engström .....	136
<b>On Dimensionality Reduction and Modelling of Pension Expenditures in Europe</b>	
Kimon Ntotsis, Marianna Papamichael, Peter Hatzopoulos, Alex Karagrigoriou .....	137
<b>An approach to nonparametric curve fitting with censored data</b>	
Jesus Orbe, Jorge Virto .....	137
<b>Stochastic comparisons of contagion risk measures in portfolios of dependent risks</b>	

<b>206</b>	<i>The 18th ASMDA International Conference (ASMDA 2019)</i>	
	Patricia Ortega-Jimenez, Miguel A. Sordo, Alfonso Suárez-Llorens.....	138
	<b>Modeling with Hyperbolic Restrictions: The Nigerian Population Dynamics</b>	
	Oyamakin Samuel Oluwafemi, Osanyintupin Olawale Dele .....	138
	<b>Model of Lifetable Evolution with Variable Drift and Cointegration</b>	
	Wojciech Otto .....	139
	<b>On the evaluation of ‘Self-perceived Age’ for Europeans and Americans</b>	
	Apostolos Papachristos, Georgia Verropoulou .....	140
	<b>Analysing the risk of bankruptcy of firms: survival analysis, competing risks and multistate models</b>	
	Francesca Pierri, Chrys Caroni.....	141
	<b>Ensemble Methods for preference structures, with relevance to the rankings’ positions</b>	
	Antonella Plaia, Simona Buscemi, and Mariangela Sciandra.....	142
	<b>Estimating age-demographic trends based on Renyi entropy</b>	
	Vasile Preda, Irina Bancescu .....	143
	<b>Latest frontiers in grouped-ordinal data dependence analysis</b>	
	Emanuela Raffinetti, Fabio Aimar .....	143
	<b>Hybrid multiple imputation for incomplete household surveys</b>	
	Humera Razzak, Christian Heumann .....	145
	<b>Weak Signals in High-dimensional Poisson Regression Models</b>	
	Orawan Reangsephet, Supranee Lisawadi, S. Ejaz Ahmed .....	145
	<b>Some New Results in Bandit and Related Problems</b>	
	Sheldon Mark Ross .....	146
	<b>Response-Adaptive Randomization: Optimizing Clinical Trials for Ethics and Efficiency</b>	
	William F. Rosenberger.....	147
	<b>Robust Bayesian analysis using multivariate classes of priors distributions</b>	
	F. Ruggeri, M. Sánchez-Sánchez, M.A. Sordo, A. Suárez-Llorens.....	147



**Heterogeneity of chronic pathology burden among elderly**

Tamara Sabgayda, Anna Edeleva, Victorya Semyonova..... 148

**Statistical Analysis of Data from Experiments Subject to Restricted Randomisation**

Aljeddani Sadiyah ..... 149

**Threshold Regression Model with Applications to the Adherence of HIV Treatment**

Takumi Saegusa, Ying Qing Chen, Mei-Ling Ting Lee ..... 150

**Health loss among the late pre-retirement and early retirement population**

Victorya Semyonova, Tamara Sabgayda..... 151

**Limiting Form for the Ergodic Distribution of a Semi-Markovian Random Walk with General Interference of Chance**

Ozlem Ardic Sevinc, Tahir Khaniyev ..... 152

**Correlation and the time interval over which the variables are measured – a non-parametric approach**

Amit Shelef, Edna Schechtman ..... 153

**Information Networks and Perturbed Markov Chains with Damping Components**

Dmitrii Silvestrov, Benard Abola, Pitos Seleka Biganda,, Sergei Silvestrov, Christopher Engstrom, John Mango Magero, Godwin A. Kakuba ..... 153

**A Novel Approach for Predicting Quality Sleep Efficiency from Wearable Device's Data**

Mayank Singh, Viranjay M. Srivastava..... 154

**Efficient Method for Lighting and Blind Control in Smart Homes to Save Energy Consumption**

Mayank Singh, Viranjay M. Srivastava..... 155

**Distributional Properties of the Percentage Change of Discrete Valued Stochastic Processes**

George-Jason Siouris and Alex Karagrigoriou ..... 156

**The Weibull model and its relationship to the Healthy years lost in a human population**

Christos H Skiadas, Charilaos Skiadas and Konstantinos N. Zafeiris ..... 156

**Branching Processes as Models of Epidemics with Vaccination Control**

Maroussia Slavtchova-Bojkova ..... 157

**Applied Meta-Analysis in Two-Class Overbooking Model**

Murati Somboon ..... 158

**Interpolation with stochastic local iterated function systems**

Soós Anna, Somogyi Ildikó ..... 159

**Acoustic evaluation of the rise-age in children's acquisition of speech sounds**

D. A. Sotiropoulos ..... 160

**Probability and Gaussian Stochastic Process applied in Engineering**

Wilker C. Sousa, Natália K. M. Galvão, Fernando F. de Souza, Regina C. B. da Fonseca ..... 161

**Performance measures in discrete supervised classification**

Ana Sousa Ferreira, Anabela Marques ..... 162

**Are there Limits for Parameter Settings in Choice-Based Conjoint Analysis?**

Winfried J. Steiner, Maren Hein, and Peter Kurz ..... 163

**Variability and the latent ageing process in life histories: Developing the statistical toolbox**

David Steinsaltz, Maria Christodoulou ..... 164

**Subset Selection of System Components for Reliability Analysis**

Eugenia Stoimenova ..... 164

**Statistical estimation in multitype branching processes with multivariate power series offspring distributions**

Ana Staneva, Vessela Stoimenova ..... 165

**SIR endemic and epidemic models in random media**

Mariya Svishchuk, Anatoliy Swishchuk, Yiqun Li ..... 166

# Assessing labour market mobility in Europe

M. Symeonaki, G. Stamatopoulou..... 167

# Describing labour market dynamics through Non Homogeneous Markov System theory

M. Symeonaki, G. Stamatopoulou..... 167

# Imputation of item non-response in Likert scales using clustering algorithms

Maria Symeonaki, Charalampos Papakonstantinou ..... 168

# A neural-network approach for predicting attitudes

Maria Symeonaki, Charalampos Papakonstantinou, Catherine Michalopoulou ..... 169

# Aging intensity order

Magdalena Szymkowiak ..... 169

# A New Method to Relate Multiblock Datasets

Essomanda Tchanda-Mangamana, Véronique Cariou, Evelyne Vigneau, Romain Glèlè Kakaï, El Mostafa Qannari ..... 170

# Robust Estimation for the Single Index Model using Pseudodistances

Aida Toma, Cristina Fulga..... 171

# Generalized Lehmann Alternative Type II Family of Distributions and its Application in Record Value Theory

Jisha Varghese, K.K. Jose..... 172

# What kind of variables can affect the JIF quartile position of a journal in Dentistry?

Pilar Valderrama-Baca, Manuel Escabias Machuca, Evaristo Jiménez-Contreras, Mariano J. Valderrama ..... 173

# Application of Markov chain process to predict the natural progression of diabetic retinopathy among adult diabetic retinopathy patients in the coastal area of South India

Senthilvel Vasudevan, Sumathi Senthilvel, Jayanthi Sureshbabu..... 173

# Multitype branching processes in random environment

<b>210</b>	<i>The 18th ASMDA International Conference (ASMDA 2019)</i>	
	Vladimir Vatutin, Elena Dyakonova, Vitali Wachtel .....	174
	<b>Research of retrieval queuing system with called applications in diffusion environment</b>	
	Viacheslav Vavilov .....	175
	<b>Clustering of variables approach in a supervised context</b>	
	Evelyne Vigneau .....	175
	<b>Probabilistic Preference Learning via the Mallows rank model: advances and case studies</b>	
	Valeria Vitelli, Øystein Sørensen, Elja Arjas and Arnoldo Frigessi .....	177
	<b>An exact polynomial in time solution of the one-dimensional bin-packing problem</b>	
	Vassily Voinov, Rashid Makarov, and Yevgeniy Voinov .....	178
	<b>Fuzzy data analysis and fuzzy directional data</b>	
	Norio Watanabe .....	178
	<b>Performance Comparison of Penalized Regression Method in Logistic Regression under High-dimensional Sparse Data with Multicollinearity</b>	
	Warangkhan Watcharasatian, Supranee Lisawadi, Benjamas Tulyanitikul .....	179
	<b>FCFS Dynamic Matching Models</b>	
	Gideon Weiss .....	179
	<b>Modeling Sporadic Event Dynamics with Markov-Modulated Hawkes Processes</b>	
	Jing Wu, Tian Zheng .....	180
	<b>Firm Technology Adoption: Optimal Timing and Employee Incentives</b>	
	Yuqian Xu .....	180
	<b>Under-five mortality in India: An application of multilevel cox proportional hazard model</b>	
	Awdhesh Yadav .....	181
	<b>The method of moments and its applications in the theory of stochastic processes</b>	

Elena Yarovaya .....	183
<b>A Markov Model for Product Line Design</b>	
Wee Meng Yeo .....	184
<b>The implications of applying alternative-supplementary measures of the unemployment rate to regions: Evidence from the European Union Labour Force Survey for Southern Europe, 2008-2015</b>	
Aggeliki Yfanti, Catherine Michalopoulou, Stelios Zachariou .....	184
<b>Improving the Understanding of Aliasing Interactions for Designed Experimentation using Regression Tree Methods</b>	
Timothy M. Young, Robert A. Breyer, Terry Liles, Alexander Petutschnigg .....	185
<b>Evaluating gender differences in Greece, 1985-2017</b>	
Konstantinos N. Zafeiris .....	186
<b>Mortality developments in Greece from the cohort perspective</b>	
Konstantinos N. Zafeiris, Anastasia Kostaki, Byron Kotzamanis .....	186
<b>TURCOSA: User-friendly, Personalized and Cloud-Based Statistical Analysis System</b>	
Gökmen Zararsız, Selçuk Korkmaz, Dinçer Göksülük, Gözde Ertürk Zararsız, Merve Başol Göksülük, Ahmet Öztürk.....	187
<b>The Rate of Growth and Fluctuations of Compound Renewal Processes</b>	
Nadiia Zinchenko .....	188
<b>Groundwater level forecasting for water resource management</b>	
Andrea Zirulia, Enrico Guastaldi, Alessio Barbagli .....	189

## Authors Index

### A

Abbad Andaloussi Samir ..... 100  
 Abdelkhalek Fatma ..... 9  
 Abdesselam Rafik..... 9  
 Abola Benard ..... 11, 35, 153  
 Adam Sobolewski Robert ..... 49, 59  
 Afanaseva Larisa ..... 11  
 Ahmadi Jafar ..... 64  
 Aimar Fabio ..... 143  
 Akbari Masoumeh..... 64  
 Alawadhi Fahimnah ..... 12  
 Albuhayri Mohammed ..... 125  
 Anandas Cheryl..... 13  
 Anastasiou A. .... 14  
 Andrade Lara ..... 93  
 Andreopoulos Panagiotis ..... 14  
 Andronov A.M ..... 15  
 Ángel Sordo Miguel..... 48  
 Anguzu Collins ..... 15  
 Anisimov Vladimir ..... 16, 23  
 Anna Soós ..... 159  
 Anthony Caruana Mark ..... 47  
 Antoni Arlette ..... 17  
 Antonov Valery ..... 18  
 Antunes Paes Neir..... 47  
 Apsemidis Anastasios..... 19  
 Araujo Pedro..... 20  
 Araujo Samuel ..... 94  
 Ardic Sevinc Ozlem ..... 152  
 Arjas Elja ..... 177  
 Arlt Josef..... 21  
 Arltová Markéta..... 21

Ascari Roberto ..... 22  
 Austin Matthew ..... 23  
 Awa Dike..... 71

### B

Baek Younghwa ..... 24  
 Bagdonavičius Vilijandas ..... 25, 26  
 Balakrishnan Narayanaswamy ..... 1  
 Balashova Daria ..... 27  
 Baldassarre Alessio ..... 59  
 Ballová Dominika ..... 27  
 Bancescu Irina..... 143  
 Baran Sándor ..... 29  
 Barbagli Alessio..... 189  
 Barone Mauro..... 29  
 Barone-Adesi Giovanni ..... 30  
 Başol Göksülük Merve ..... 187  
 Bastianin Andrea..... 31  
 Batista Carvalho João..... 47  
 Behtari Behnam ..... 32  
 Benšić Mirta..... 33  
 Bersimis Fragkiskos G..... 14  
 Bey Siham ..... 34  
 Bieniek Mariusz ..... 35  
 Blanchet Jose ..... 108  
 Bodnar Olha..... 130  
 Bolla Marianna ..... 9  
 Booth Heather ..... 36  
 Bougeard Stéphanie..... 39  
 Boussaha Aicha ..... 40  
 Bozeman James R. .... 41  
 Brettschneider J.A. .... 55  
 Breyer Robert A. .... 185

Brown Mark .....	1
Bry Xavier .....	39
Bulinskaya Ekaterina .....	42
Burkschat Marco .....	35
Buscemi Simona .....	142

## C

Canhanga Betuel .....	45, 134
Carette Philippe .....	95
Cariou Véronique .....	119, 170
Carlsson Linus .....	24
Caroni Chrys .....	46, 141
Castaño Antonia .....	48
Castro Juan Eloy Ruiz .....	49
Casula Laura .....	49, 50
Charalampi Anastasia .....	51
Chernykh German .....	52, 74
Choon Chia Ngee .....	53
Christodoulou M.D. ....	55
Christodoulou Maria .....	164
Chukova Stefanka .....	56
Coelho Edvigés .....	41
Conversano Claudio .....	59
Cordeiro M.T.A. ....	88
Crippa Franca .....	56
Cristina Canepa Elena .....	44
Cui Qi .....	36

## D

D'Ambrosio Antonio .....	59
D'Amico Guglielmo ..	49, 59, 60, 61, 62, 63
da Fonseca Regina C.B. ....	161
Dai Hongsheng .....	58
Dalinger I.M. ....	15
David Chikezie .....	71
Davydov Roman .....	18
de Almeida Botega Laura .....	37

de Souza Fernando F. ....	161
del Mar Rueda-García María ....	114
Dembińska Anna .....	64, 65
Derquenne Christian .....	66
Desbois Dominique .....	67
Devolder Pierre .....	68, 70
Dhorne Thierry .....	17
Di Lorenzo Emilia .....	69
Di Prignano Ernesto Volpe .....	62
Diakite Keivan .....	70
Dimitrakos Theodosios D. ....	72, 113
Dimotikalis Yiannis .....	73
Dmitrieva Ludmila .....	52, 74
Duarte Denise .....	75
Dubinin N.E. ....	76
Dyakonova Elena .....	174

## E

Edeleva Anna .....	148
Ediev Dalkhat M. ....	76
Ejaz Ahmed S. ....	145
Elliott Robert J. ....	2
Engström Christopher ....	11, 15, 35, 77, 125, 136
Eric Otuonye .....	71
Ertürk Zararsız Gözde .....	187
Escabias Machuca Manuel .....	173

## F

Faiza Asif .....	79
Fayyaz Saeed .....	81
Fazekas István .....	82
Fedorov Valerii .....	83
Fiala Tomas .....	83
Filus Jerzy .....	84
Filus Lidia .....	84
Fotopoulos S.B. ....	85
Freeman Jim .....	86

Frigessi Arnoldo .....	177
Fulga Cristina.....	171
Fung Stanley Zel Tsz .....	54

## G

G. Udomboso Christopher .....	121
Gadrich Tamar .....	86
Galvão Natália K.M. ....	161
García Jesús E. ....	87, 88
Garnie-Zarli Evelynne .....	100
Gatto Gwenaël.....	90
Giordano Francesco .....	90
Girardin Valerie.....	3
Girardin Valrie .....	91
Glèlè Kakaï Romain .....	170
Göksülük Dinçer .....	187
González del Pozo Raquel .....	89
González-López V.A.....	87, 88
Gormley Stephen .....	23
Gorncy Agnieszka.....	64
Grabski Franciszek .....	92
Grishunina Svetlana .....	11
Groll Andreas .....	92
Guambe Calisto.....	45, 134
Guastaldi Enrico .....	189
Guedes Gilvan.....	20, 75, 93, 94
Guerry Marie-Anne .....	95
Guessoum Zohra .....	34

## H

Hadjiliadis Olympia .....	90, 98
Haghiri Maryam .....	96
Haji Babak.....	96
Hamedović Safet.....	33
Hamel Andreas .....	98
Hamieh Tayssir .....	100
Hamrani Farida .....	97
Hassiba Benseradj.....	97

Hatzopoulos P.....	14, 99
Hatzopoulos Peter .....	137
Hayek Ali.....	100
Hein Maren.....	163
Heiner Jonas .....	92
Helena Spyrides Maria .....	93
Heumann Christian .....	145
Hon Filip .....	101
Hruschka Harald .....	101
Huang Chao .....	58
Huber-Carol Catherine .....	102

## I

Ilori Olamide O.....	121
Iorio Carmela .....	103
Ivanova Alla .....	104

## J

Jandhyala V.K.....	85
Janssen Jacques .....	62
Jennison Christopher .....	105
Jensen Christian .....	105
Jiménez-Contreras Evaristo.....	173
Jin Lu.....	106
Jizba Petr .....	106
Johnson Miriam J. ....	58
Jose K.K.....	107, 172
Joseph Jeena.....	107
Jrada M Es-salih Ben .....	78
Jung Kyungsik .....	24

## K

Kai Choy Siu .....	54
Kakuba Godwin.....	11, 35, 122
Kakuba Godwin A.....	153
Kang Yang .....	108
Karagrigoriou A. ....	14
Karagrigoriou Alex .....	137, 156



Karamatsoukis Constantinos C. . 72, 113	
Karanja Elvis..... 108, 109	
Katehakis Michael N. .... 3	
Kaul A. .... 85	
Keikara Muhumuza Asaph ..... 122	
Khadidja Djaballah ..... 78	
Khalidi Khaled ..... 127	
Khaniyev Tahir ..... 152	
Khmaladze Estate V. .... 43	
Khokhlov Yury ..... 110	
Khraibani Zaher..... 100	
Kim Hoseok..... 24	
Kim Juhan ..... 116	
Kleeva Daria..... 74	
Kolias Pavlos ..... 111	
Korkmaz Selçuk..... 187	
Korolev Victor ..... 110	
Kössler W..... 117	
Kostaki Anastasia ..... 186	
Kostner Daniel ..... 98	
Kotzamanis Byron ..... 186	
Kovalenko Anatoly ..... 18	
Krzyśko Mirosław ..... 111	
Kubilius Kęstutis..... 113	
Kumar Mukhopadhyay Barun .. 131	
Kuperin Yuri ..... 52, 74	
Kurz Peter ..... 163	
Kyriakidis Epaminondas G. . 72, 113	

## L

Labenne Amaury..... 119	
Laeven Roger J.A. .... 43	
Lam Mo Yee ..... 51	
Langhamrova Jitka ..... 83, 101	
Lazarova Meglena D..... 115	
Lebedev Eugene O. .... 119	
Lee Taerim ..... 116	

Lefèvre Claude ..... 4	
Lenz H.J..... 117	
Leonilde Carli Lucia ..... 56	
Levulienė Rūta ..... 26	
Li Shuangning..... 117	
Li Nan ..... 118	
Li Yiqun ..... 166	
Liles Terry ..... 185	
Limnios Nikolaos ..... 126	
Lisawadi Supranee ..... 145, 179	
Liu Ruiping ..... 7	
Livinska Hanna V. .... 119	
Ildikó Somogyi..... 159	
Llobell Fabien..... 119	
Loschi Rosangela..... 20	
Luis García-Lapresta José ..... 89	
Lundengård Karl..... 122	
Lykou Rodi ..... 123	
Lyu S. .... 85	

## M

M. Di Brisco Agnese ..... 68	
Maina Naomi ..... 108	
Majewska Justyna..... 123	
Makarov Rashid ..... 178	
Malinowski Marek T..... 124	
Malyarenko Anatoliy 125, 132, 133, 136	
Manca Raimondo..... 62	
Mango John ..... 11, 35, 122, 153	
Mango Magero John ..... 11, 35, 153	
Marcoulides Katerina M..... 126	
María Lara-Porras Ana ..... 114	
Mark Ross Sheldon ..... 146	
Marques Anabela..... 162	
Marrime Alex ..... 134	
Masala Giovanni ..... 49, 50, 59	
Mashhadi Heidar Ashraf ..... 80	

Mavridoglou G. ....	14
Maya <b>Brenda Ivette Garcia</b> .....	126
McClellan Sally.....	7
Meddahi Samia .....	127
Medžiūnas Aidas .....	113
Meilijson Isaac .....	128
Meng Yeo Wee .....	184
Menni Nassira .....	128
Mercado Londoño S.L. ....	88
Mi Hong .....	118
Michalopoulou Catherine .	51, 169, 184
Migliorati Sonia.....	22, 68
Miguel Bravo Jorge .....	41
Minkova Leda D. ....	56, 115, 129
Molchanov Stanislav .....	27
Molina-Muñoz David .....	114
Morand Elisabeth.....	129
Mostafa Qannari El .....	119, 170
Moutti Maria .....	14
Murara Jean-Paul.....	132, 133
Murthy Karthyek.....	108

## N

Nalule Muhumuza Rebecca.....	130
Nankina Loy .....	134
Nhangumbe Clarinda .....	134
Ni Ying .45, 106, 125, 132, 133, 136	
Niang Ndèye .....	7, 39
Niglio Marcella .....	90
Nikitin Yakov .....	135
Nikolov Nikolay I. ....	65
Nohrouzian Hossein .....	136
Noor-ul-Amin Muhammad .....	79
Noronha Kenya .....	93, 94
Noszály Csaba .....	82
Nsubuga Rebecca.....	130
Ntotsis Kimon .....	137

Nwe Aye Tin.....	24
Nzabanita Joseph .....	130

## O

Okrasa Włodzimierz .....	111
Olawale Dele Osanyintupin .....	138
Ongaro Andrea .....	22
Orbe Jesus .....	137
Ortega-Jimenez Patricia .....	138
Osakwe Carlton.....	2
Otto Wojciech.....	139
Oulidi Abderrahim .....	70
Öztürk Ahmet .....	187

## P

Pandolfo Giuseppe .....	103
Papachristos Apostolos .....	140
Papadopoulou Aleka .....	111
Papakonstantinou Charalampos .....	168, 169
Papamichael Marianna .....	137
Pechholdová Markéta .....	57
Petkevičius Linas .....	25
Petroni Filippo ....	49, 59, 60, 61, 62
Petutschnigg Alexander .....	185
Pierri Francesca .....	141
Pigueiras Gema .....	48
Pirvu Traian A. ....	44
Pisani Stefano .....	29
Pisati Matteo M. ....	30
Piscopo Gabriella .....	69
Plaia Antonella .....	142
Preda Vasile .....	143
Prokš Martin .....	106
Psarakis Stelios .....	19

## Q

Qing Chen Ying.....	150
---------------------	-----

**R**

Raffinetti Emanuela .....	31, 143
Ragozin Ilya .....	135
Ramalho Guedes Gilvan .....	37
Rancic Milica .....	122, 132
Razzak Humera .....	145
Reangsephet Orawan .....	145
Regnault Philippe .....	61
Restaino Marialuisa .....	58
Ribeiro Paula .....	94
Ribeiro Rodrigo .....	75
Richardson Clive .....	51
Rosenberger William F. ....	147
Ruggeri F. ....	147
Ryazantsev Sergey .....	104
Rychlik Tomasz .....	35

**S**

Sabgayda Tamara .....	148, 151
Sabo Kristian .....	33
Sadiah Aljeddani .....	149
Saegusa Takumi .....	150
Sagianou A. ....	99
Sakurai Yuji .....	106
Sala Carlo .....	30
Samo June .....	108
Samuel Oluwafemi Oyamakin ..	138
Sánchez-Sánchez M. ....	147
Saporta Gilbert .....	7
Schauberger Gunther .....	92
Schechtman Edna .....	153
Sciandra Mariangela .....	142
Scocchera Stefania .....	61, 63
Seleka Biganda Pitos .....	11, 35, 153
Šeliga Adam .....	27
Semyonova Victorya .....	148, 151
Senoussi Rachid .....	91
Senthilvel Sumathi .....	173

Shelef Amit .....	153
Shing Chan Ping .....	51
Shoshanna Procaccia Rossella ....	56
Sibillo Marile .....	69
Silva Claudio .....	93
Silva Pollyane .....	93
Silvestrov Dmitrii .....	8, 153
Silvestrov Sergei .....	11, 15, 35, 122, 125, 130, 132, 133, 153
Singh Mayank .....	154, 155
Siouris George-Jason .....	156
Skiadas Charilaos .....	156
Skiadas Christos H. ....	4, 156
Slavtchova-Bojkova Maroussia. ....	157
Somboon Murati .....	158
Sordo M.A. ....	147
Sordo Miguel A. ....	138
Sørensen Øystein .....	177
Sotiropoulos D.A. ....	160
Sotoudeh Farkosh Reza .....	80
Sousa Ferreira Ana .....	162
Sousa Wilker C. ....	161
Spingola Andrea .....	29
Spizzichino Fabio L. ....	5
Srivastava Viran Jay M. ....	154, 155
Stamatopoulou G. ....	167
Staneva Ana .....	165
Steiner Winfried J. ....	163
Steinsaltz D. ....	55
Steinsaltz David .....	164
Stoimenova Eugenia .....	65, 164
Stoimenova Vessela .....	165
Storchi Lorian .....	61, 63
Suárez-Llorens Alfonso .....	138
Suárez-Llorens A. ....	147
Suleiman Samya .....	122
Sulemana Hisham .....	122
Sureshbabu Jayanthi .....	173

Svačinová Kornélia .....	57
Svishchuk Mariya .....	166
Swishchuk Anatoliy .....	166
Symeonaki M. ....	167
Symeonaki Maria .....	168, 169
Szymkowiak Magdalena.....	169

**T**

Tabaja Nabil .....	100
Tatachak Abdelkader .....	34, 128
Tchandao-Mangamana Essomanda .....	170
Tewolde Finnan .....	125
Ting Lee Mei-Ling .....	150
Tizzano Roberto .....	69
Toma Aida .....	171
Toufaily Joumana .....	100
Tragaki Alexandra .....	14
Trinchera Laura.....	126
Trzpiot Grazyna.....	123
Tsaklidis George.....	123
Tulyanitikul Benjamas .....	179

**U**

Uhrmeister Jörn .....	92
Umut Can Sami .....	43
Uzonyi Noémi .....	82

**V**

Valderrama Mariano J. ....	173
Valderrama-Baca Pilar .....	173
Varghese Jisha .....	172
Vassiliou P.-C.G. ....	6
Vasudevan Senthilvel .....	173
Vatutin Vladimir.....	174
Vavilov Viacheslav.....	175
Verron Thomas .....	39
Verropoulou Georgia .....	140

Viegas Andrade Mônica .....	37
Vigneau Evelyne.....	119, 170, 175
Virto Jorge .....	137
Vitelli Valeria .....	177
Vitiello Autilia .....	103
Voinov Vassily .....	6, 178
Voinov Yevgeniy.....	178
Vrabcová Jana .....	57

**W**

Wachtel Vitali .....	174
Wang Huiwen .....	7
Watanabe Norio .....	178
Watcharasatian Warangkhan .	179
Weiss Gideon.....	179
Wen Yijin .....	86
Wołyński Waldemar.....	111
Wu Jing.....	180

**X**

Xu Yuqian.....	180
----------------	-----

**Y**

Yackiv I.V. ....	15
Yadav Awdhesh.....	181
Yarovaya Elena .....	27, 183
Yfanti Aggeliki .....	184
Young Timothy M. ....	185

**Z**

Zachariou Stelios.....	184
Zafeiris Konstantinos N. ...	156, 186
Zararsız Gökmen .....	187
Zeddouk Fadoua .....	68
Zemlyanova Elena .....	104
Zenga Mariangela .....	49, 56
Zhang Fan .....	108
Zhang Jiahui .....	125

Zheng Tian .....	180	Zirulia Andrea .....	189
Zinchenko Nadiia .....	188	Zohra Guessoum.....	97